The Optimal Degree of Flexibility in Policy Rules

Guillaume Sublet

University of Montreal [†]

July 21, 2026

The design of a rule involves a trade-off between discipline and flexibility. I use global Lagrangian methods to obtain necessary as well as sufficient conditions that completely characterize optimal rules in a general model of delegation in which marginal deterrence is socially costly (i.e., burns money). The optimal degree of flexibility depends on the agent's bias, the principal's aversion to burning money, and the thickness of the tail of the distribution of shocks. If the tail is thin relative to the bias and the aversion to burning money, an optimal rule imposes discipline by restricting the set of implementable policies. If the tail is relatively thick, instead, an optimal rule imposes discipline through marginal deterrence, which preserves flexibility at the cost of burning money. Applications to the design of fiscal rules, the design of trade agreements, and the regulation of a monopolist illustrate the key trade-offs and suggest avenues for reform.

KEYWORDS: Policy Rule, Delegation, Mechanism Design, Money Burning, Flexibility, Price vs. Quantity, Fiscal Rule, Trade Agreement, Monopoly Regulation

JEL CLASSIFICATION: D02, D82, D86, E62, F13, L51

[†] Email: guillaume.sublet@umontreal.ca.

This paper expands the theory developed in a previous paper circulated under the title "The Optimal Degree of Discretion in Fiscal Policy." I thank Manuel Amador, Marina Halac, Pierre Yared, and referees for comments and suggestions that greatly improved the paper. I also thank Marco Bassetto, Anmol Bhandari, Roel Beetsma, Job Boerma, Rui Castro, Louphou Coulibaly, Alessandro Dovis, Mike Golosov, Tim Kehoe, Rishabh Kirpalani, Ellen McGrattan, Thomas Phelan, Todd Schoellman, Bob Staiger, Pierre-Olivier Weill and the audiences of academic seminars and conferences for helpful suggestions and comments. I am responsible for any shortcomings of this paper.

Introduction

The design of a policy rule involves a central trade-off between imposing discipline to act in the interest of society and granting flexibility to respond to shocks. Should a policy rule balance discipline and flexibility with a graduated schedule of penalties, a restriction on the set of implementable policies, or a combination of the two?

The literature provides a sharp answer for two polar cases. If the societal cost of levying a penalty on policy-makers is negligible, then the optimal rule features a graduated schedule of penalties resembling a Pigouvian tax (Walsh (1995)). In contrast, if penalties are joint—in the sense that the cost of penalties affects society and the policy-maker symmetrically—then the optimal rule necessarily features bang-bang incentives (Halac and Yared (2022a)). In that case, under a condition on the distribution of shocks, the optimal rule restricts the set of implementable policies to an interval (Athey, Atkeson, and Kehoe (2005), Amador, Werning, and Angeletos (2006)). Little is known, however, about the optimality of rules featuring a graduated schedule of penalties within a restricted set of implementable policies.

In this paper, I study the optimal design of rules in a general principal-agent model of the trade-off between discipline and flexibility as in Amador and Bagwell (2013). The rule is enforced by penalties whose burden may be borne asymmetrically by the policy-makers and the society. The focus is on characterizing optimal rules for degrees of asymmetry between the two polar cases. The main insights from the two polar cases extend to intermediate asymmetry, but only under more stringent conditions. Under a complementary set of conditions, the optimal rule instead combines the two instruments, imposing a graduated schedule of penalties within a restricted set of implementable policies.

The main contribution of this paper is the development of tools that facilitate the design of optimal rules. While the existing literature has developed powerful tools—namely, sufficient optimality conditions—to verify that a candidate rule is globally optimal, the task of constructing such a rule remains challenging. I extend the applicability of global Lagrangian methods used in Amador, Werning, and Angeletos (2006) and Amador and Bagwell (2013, 2020) to derive necessary and sufficient optimality conditions that provide a complete characterization of solutions to delegation problems with the possibility of burning money.

This contribution builds on a line of research studying the trade-off between discipline and flexibility in the design of a rule (Amador and Bagwell (2013) and Halac and Yared (2020a)). Consider a principal-agent model in which a policy must be chosen in response to a shock. The principal aims to delegate the policy-making authority to the agent before observing the

realization of the shock. The agent's objective, however, is biased relative to that of the principal. As a result, a well-designed policy rule must balance the need to discipline a biased agent with the need for flexibility to respond to shocks. Complicating matters, there is a trade-off between the need for discipline and flexibility because the realization of the shock is not contractible (e.g., it may be private information to the agent), so the rule cannot condition the penalty on the shock itself. Hence a policy rule provides incentives to the agent by specifying a penalty as a function of the policy chosen. The task of optimally designing a policy rule maps to a delegation problem with money burning, where money burning refers to the cost for the principal of meting out a penalty on the agent. Money burning is symmetric if the penalties are joint for the principal and the agent and asymmetric otherwise.

Characterizing optimal rules is challenging because the standard first-order approach does not apply. Unlike local incentive compatibility constraints, the non-negativity constraints on money burning cannot be easily summarized in the objective function. Global Lagrangian methods conveniently account for the non-negativity constraints on money burning to deliver sufficient optimality conditions (Amador and Bagwell (2013, 2020), Amador, Bagwell, and Carpizo (2026)). Given a guess of a rule and a Lagrange multiplier function, the sufficient optimality conditions verify the optimality of the rule (Amador and Bagwell (2013, 2020)). The remaining challenge is to guess the optimal rule and the associated Lagrange multiplier. This is where necessary optimality conditions are useful. Powerful tools have been developed to obtain necessary optimality conditions for rules enforced by symmetric money burning—perturbation methods (Athey, Atkeson, and Kehoe (2005), Halac and Yared (2022a)), majorization and extreme-point techniques (Kleiner, Moldovanu, and Strack (2021)), and an equivalence with Bayesian persuasion (Kolotilin and Zapechelnyuk (2025)). When money burning is not symmetric, however, the optimal rule need not be an extreme point.

The global Lagrangian methods, as used in this paper, apply to both symmetric and asymmetric money burning and allow for rules with discontinuities. The challenge in obtaining necessary optimality conditions for any rule is to account for the shadow cost on the money-burning constraint. I use an approximation argument to obtain a Lagrange multiplier functional—which captures the shadow cost on the money-burning constraint—amenable to using standard tools of optimization. The resulting necessary optimality conditions, which must hold for any optimal rule, partially determine whether the optimal compromise between discipline and flexibility takes the form of marginal deterrence, bunching, or discretion. The conditions depend on three key features of the model: the bias governs the need for discipline; the thickness of the tail of the distribution of shocks governs the need for flexibility; and the degree of asymmetry in the cost

of burning money governs the principal's aversion to burning money.

I show that marginal deterrence can be part of an optimal rule only if the tail of the distribution of shocks is thick relative to the principal's aversion to burning money and the agent's bias. The intuition is that, because allowing for a flexible, yet disciplined, response to shocks burns money, it is only worthwhile if the need for flexibility, governed by the thickness of the tail of the distribution of shocks, is sufficient. First, I obtain a closed-form expression for the wedge in the Euler equation of the agent that optimally balances the benefit of discipline with the incentive cost of flexibility. This characterization shows that, in the limit case of symmetric money burning, the use of bang-bang incentives is a generic property of an optimal rule. Second, for marginal deterrence to be compatible with incentives, the tail of the distribution of shocks must be sufficiently thick—that is, the slope of the inverse hazard rate lies above a threshold set by the agent's bias and the principal's aversion to burning money—so that the policy schedule is monotonic. A failure of the monotonicity condition, for a rule that burns money, implies that bunching is optimal instead.

Bunching can be part of an optimal rule for two distinct reasons. The first one is standard. Bunching is optimal when the tail of the distribution of shocks is sufficiently thin, so that the incentive cost of flexibility outweighs the benefit. This echoes the optimality of bang-bang incentives for symmetric money burning and, for asymmetric money burning, corresponds to ironing a non-monotonic policy plan. In addition, bunching may be optimal due to an exemption from marginal deterrence. That is, even when the policy plan consistent with marginal deterrence is monotonic, it may be optimal to forgo marginal deterrence below a threshold to economize on money burning. The resulting kink in the penalty schedule causes bunching. Despite these distinct economic motives for bunching, the global Lagrangian methods deliver a unified necessary optimality condition for bunching in any optimal rule, including rules with discontinuities. Lastly, I show that granting discretion (i.e., the absence of discipline) can be optimal only if the bias is sufficiently small relative to the incentive cost of discipline. I obtain this result by generalizing the standard optimality condition for discretion in interval delegation to allow for asymmetry in the cost of burning money.

These necessary conditions provide the building blocks of globally optimal rules. I show that the necessary conditions for marginal deterrence, bunching, and discretion are jointly sufficient. To build a globally optimal rule, it suffices to piece together a non-decreasing policy plan that is locally optimal within each piece of the partition of the shocks over which there is marginal deterrence, bunching, or discretion, and such that the Lagrange multiplier is non-decreasing across the pieces of the partition. I use this construction method to completely characterize

the conditions under which each of two commonly used classes of rules is optimal. The first class, interval delegation, does not burn money. The second class consists of rules enforced by a graduated schedule of penalties, which achieve a finer balance between discipline and flexibility at the cost of burning money above a threshold. To economize on money burning, it may be optimal to grant an exemption from marginal deterrence below a threshold. Which class is optimal, and how lenient the rule is, depend on the same three features of the environment that govern the necessary conditions: the bias, thickness of the tail of the distribution of shocks, and the degree of asymmetry in the cost of burning money. Where the tail of the distribution of shocks is thick, an optimal rule relies on marginal deterrence, preserving flexibility at the cost of burning money. Where the tail is thin, an optimal rule disciplines the agent by restricting the set of implementable policies.

I illustrate the reach of the tools in three applications. Across three disparate settings, the necessary optimality conditions translate observed institutions into testable restrictions on fundamentals, and the sufficient optimality conditions construct optimal rules and suggest avenues for reform. In the design of a fiscal rule, the rule must discipline a deficit-biased government while preserving flexibility to respond to shocks to the fiscal needs. With asymmetric money burning, the optimal rule ranges from a simple cap on deficits to a graduated schedule of penalties as the need for flexibility—governed by the thickness of the tail of the distribution of shocks to fiscal needs—grows relative to the need for discipline—governed by the deficit bias. In the design of a trade agreement, the cost of burning money reflects the inefficiency with which an importing country compensates its trading partners for protection. The theory shows that the WTO's tariff bindings, tariff overhang, and safeguards can all be features of an optimal agreement. According to the theory, however, making the penalty for invoking a safeguard depend on the size of the tariff in excess of the binding—in addition to the duration of its prior use—would improve upon the current design. In the regulation of a monopolist with unknown costs, the cost of burning money reflects the degree of regulatory capture. As capture intensifies, the optimal regulation moves away from marginal-cost pricing toward bang-bang incentives, converging at sufficiently high capture to no discipline at all—the monopolist's flexible price. The thickness of the right tail of the cost distribution governs whether the regulated price schedule steepens or flattens along the way. The leniency on fines prescribed by the theory is consistent with the contrast between U.S. and European antitrust doctrine on excessive pricing.

Related literature

More broadly, this paper relates to the literature on delegation and the design of rules that discipline a policy-making authority to act in the interest of society. Relative to the literature on delegation without money burning (Holmström (1977), Melumad and Shibano (1991), Alonso and Matouschek (2008)), allowing for money burning relaxes the constraint set and enlarges the set of implementable rules (Amador, Werning, and Angeletos (2006), Ambrus and Egorov (2017), Amador and Bagwell (2013, 2020, 2022), and Halac and Yared (2014, 2018, 2020a, 2020b, 2022a, 2022b, 2025)).

The paper is most closely related to the strand of this literature in which the optimal rule burns money. Money burning may arise because enforcement is limited (Halac and Yared (2022a, 2025)). Alternatively the optimality of rules that provide incentives with marginal deterrence at the cost of burning money has been studied in Mookherjee and Png (1994) when the bias of the agent is extreme, Ambrus and Egorov (2017) in a linear-quadratic environment, and Amador and Bagwell (2020) who provide sufficient optimality conditions in a general principal-agent model. The paper contributes to the literature along two dimensions: a methodological one, by providing necessary and sufficient optimality conditions that completely characterize optimal rules, including rules with discontinuities (see Amador, Bagwell, and Carpizo (2026) for sufficient optimality conditions), and a substantive one, by characterizing optimal rules across the full range of asymmetry in the cost of burning money in a general principal-agent model of the trade-off between discipline and flexibility.

This paper relates to the literature on bunching. Although money burning adds a second source of bunching beyond the standard ironing procedure, convexity of the problem again delivers a characterization. Prior work invokes convexity for incentive reasons. Generalized ironing convexifies an exogenous virtual value in the unidimensional setting (Nöldeke and Samuelson (2007), Toikka (2011)), while incentive compatibility requires convexity of the indirect utility in the multidimensional setting (Rochet and Choné (1998), Boerma, Tsyvinski, and Zimin (2025)). Global Lagrangian methods instead exploit the convexity of the region below the graph of the principal's value function, obtaining the Lagrange multiplier as a supporting hyperplane. This Lagrange multiplier function endogenizes the virtual value to account for the shadow cost of the money-burning constraint, yielding a Diamond (1998)-type ABC formula in which the tail thickness of the distribution of shocks sets the progressivity of the incentive scheme. As in ironing, bunching arises to preserve incentive compatibility. In addition, bunching may be optimal even absent any need for ironing, emerging instead from an exemption from marginal deterrence that economizes on money burning.

Although this paper restricts attention to deterministic rules, the theory applies to characterize optimal stochastic mechanisms in environments with mean-variance preferences. Stochastic mechanisms can improve upon deterministic ones in related delegation and communication problems (Goltsman, Hörner, Pavlov, and Squintani (2009), Kováč and Mylovanov (2009), Kartik, Kleiner, and Van Weelden (2021)). Hence, the paper also relates to work on the benefit of opacity in the prescription of rules, which can be desirable under reputational considerations (Dovis and Kirpalani (2021)).

Because a graduated schedule of penalties resembles a price instrument and a restriction of the choice set resembles a quantity instrument, this paper also contributes to the literature on the choice of regulatory instrument. In an influential contribution, Weitzman (1974) shows that the relative curvature of the benefit and cost functions of economic activity determines whether fixing the price or fixing the quantity is preferable.[2] Perhaps surprisingly, the distribution of shocks does not affect the ranking of these two simple instruments—it only affects the intensity of the preference for one over the other. Allowing for a richer set of instruments, as I do in this paper, reveals the role of the distribution of shocks for the choice of regulatory instrument: a thick tail of the distribution of shocks calls for flexibility, which favors marginal deterrence with price regulation.[3]

1 Model

This section introduces a general principal-agent model of the trade-off between discipline and flexibility in the design of a rule as in Amador and Bagwell (2013). The model nests models of the design of incentive contracts for policy-makers—in which case the penalty is borne privately by the policy-maker as in Walsh (1995)—and models of delegation with symmetric money burning—in which case penalties are joint as in the model of Halac and Yared (2022a).

The welfare of the principal $U^P(\theta, \pi)$ depends on a *shock*, denoted $\theta \in [\underline{\theta}, \bar{\theta}]$, and a policy $\pi \in [\underline{\pi}, \bar{\pi}]$. The welfare of the principal is twice continuously differentiable in both its arguments, strictly concave in π, and it satisfies the single-crossing condition $U^P_{\theta\pi} > 0$. The shock follows a distribution F, with continuously differentiable density f, whose support is the interval $[\underline{\theta}, \bar{\theta}]$

[2]For instance, the consensus in environmental economics is that the marginal benefit curve of abatement is relatively flat, whereas the marginal cost curve is steep, which according to Weitzman (1974) favors price regulation over quantity regulation (see McKibbin and Wilcoxen (2002)).

[3]The design of rules focuses on the trade-off between discipline and discretion by abstracting from subsidization across types. See Akbarpour, Dworczak, and Kominers (2024) for a mechanism design approach with a concern for redistribution.

where $0 < \underline{\theta} < \bar{\theta} < \infty$.

The welfare of the agent is $U^A(\theta, \pi) = \theta\pi + b(\pi)$, with b twice continuously differentiable and $b''(\pi) < 0$ for $\pi \in [\underline{\pi}, \bar{\pi}]$. The outcome under discretion, denoted $\pi_f(\theta)$, maximizes $U^A(\theta, \pi)$ and it is assumed to be interior. Let ν denote the difference between the agent and the principal's welfare, $\nu(\theta, \pi) = U^A(\theta, \pi) - U^P(\theta, \pi)$. Let $\kappa \equiv \inf_{\theta \in [\underline{\theta}, \bar{\theta}], \pi \in [\underline{\pi}, \bar{\pi}]} \{U^P_{\pi\pi}(\theta, \pi)/U^A_{\pi\pi}(\theta, \pi)\}$. Unless specified otherwise, the principal's welfare is normalized so that $\kappa = 1$.

A *rule* (π, τ) is a policy function $\pi : [\underline{\theta}, \bar{\theta}] \to \mathbb{R}$ and a money burning schedule $\tau : [\underline{\theta}, \bar{\theta}] \to \mathbb{R}$. Money burning refers to an action that penalizes both the agent and the principal. Let $\rho \geq 0$ denote the cost of money burning for the principal relative to that for the agent. A rule is *optimal* if it maximizes the objective of the principal

$$\int_{\underline{\theta}}^{\bar{\theta}} \left[U^P(\theta, \pi(\theta)) - \rho\tau(\theta)\right] dF(\theta) \tag{1}$$

subject to the incentive compatibility constraint

$$U^A(\theta, \pi(\theta)) - \tau(\theta) \geq U^A(\theta, \pi(\hat{\theta})) - \tau(\hat{\theta}) \quad \text{for } \theta, \hat{\theta} \in [\underline{\theta}, \bar{\theta}], \tag{2}$$

and the money-burning constraint

$$\tau(\theta) \geq 0 \quad \text{for } \theta \in [\underline{\theta}, \bar{\theta}]. \tag{3}$$

A rule is *admissible* if it satisfies the constraints (2) and (3). The incentive compatibility constraints (2) are standard. In the applications, it follows from the non-contractible nature of the shocks to the benefit of the policy (e.g., the realization of θ is private information of the agent). In the applications, the money-burning constraint arises naturally given that the rule provides incentives with penalties that are joint for the principal and the agent to a degree parametrized by $\rho > 0$. One way to penalize a risk-averse agent is to add noise to the prescription of the rule. For a principal and an agent who both have quadratic mean-variance preferences, money burning can be interpreted as adding noise to the prescription of the rule (see Kartik, Kleiner, and Van Weelden (2021) for a general treatment of stochastic mechanisms). In this case, the money-burning constraint captures the restriction that the variance of the noise is positive. Without the money-burning constraint, the principal would have an incentive to provide unbounded rewards to the agent. Note, however, that the lower bound at zero is a normalization. A different lower bound would only affect an optimal rule by shifting the money-burning schedule uniformly.

The money-burning constraint distinguishes the design of an optimal rule from the design of mechanisms with and without transfers. Importantly, the money-burning constraint rules out subsidization across different realizations of θ, unlike models of the trade-off between equity and

efficiency in which $\int_{\underline{\theta}}^{\bar{\theta}} \tau(\theta) dF(\theta) = 0$ instead (as in, for instance, Atkeson and Lucas (1992), Galperti (2015)). The money-burning constraint also differs from delegation models in which $\tau(\theta) = 0$ for $\theta \in [\underline{\theta}, \bar{\theta}]$ (as in, for instance, Holmström (1977) and Halac and Yared (2020a)). Note, however, that if the optimal rule does not burn money, then it also solves the mechanism design problem with $\tau(\theta) = 0$ for $\theta \in [\underline{\theta}, \bar{\theta}]$ instead of the money-burning constraint.[4]

In considering the full range of the degrees of asymmetry in the cost of enforcement $\rho \in [0, 1]$ (equivalently $\rho \in [0, \kappa]$ without the normalization of the relative concavity in payoffs), this paper bridges the gap between two literatures. For $\rho = 0$, the principal is immune to the penalties on the agent, which is a limit case of the design of mechanisms with transfers (as in Walsh (1995), and more recently Clayton and Schaab (2025)).[5] For $\rho = 1$, penalties symmetrically affect the agent and the principal (Halac and Yared (2022a)). More precisely, money burning is *symmetric* if the relative curvature of the objectives of the principal and the agent is constant (i.e., $U^P_{\pi\pi}(\theta, \pi) = \kappa U^A_{\pi\pi}(\theta, \pi)$ for all θ, π) and $\rho = \kappa$.

The Ramsey and the expected Ramsey rules are useful benchmarks. The *Ramsey rule* achieves the highest welfare for the principal in an environment in which θ is contractible information; that is the Ramsey policy $\pi^P(\theta)$ solves $\max_{\pi \in [\underline{\pi}, \bar{\pi}]} U^P(\theta, \pi)$ and, because there is no need to burn money to provide incentives when information is contractible, $\tau^P(\theta) = 0$ for $\theta \in [\underline{\theta}, \bar{\theta}]$. The Ramsey policy is strictly increasing (by the single-crossing condition) and assumed to be in the interior of $[\underline{\pi}, \bar{\pi}]$. The Ramsey rule is not admissible because it is not incentive compatible (unless the agent is not biased, in which case $\pi_f(\theta) = \pi^P(\theta)$). The *expected Ramsey rule* is the rule with the constant policy that achieves the highest welfare for the principal and does not burn money. The constant policy π^{ER} solves $\max_{\pi \in [\underline{\pi}, \bar{\pi}]} \int_{\underline{\theta}}^{\bar{\theta}} U^P(\theta, \pi) dF(\theta)$. While the expected Ramsey rule is admissible, it need not be optimal.

The design of a rule in the polar case in which the principal is immune to the penalty on the agent—i.e., $\rho = 0$—can be solved from first principles because it implements the Ramsey policy (hence, the main analysis presumes $\rho > 0$ unless specified otherwise). Intuitively, without concern for the societal cost of burning money, the money burning schedule can be designed with the sole purpose of implementing the Ramsey policy. Formally, let the function τ be given by the money burning schedule (4) that implements the Ramsey policy with $\tau(\underline{\theta})$ such that the money-burning constraint is satisfied. Because $\rho = 0$, this rule achieves the welfare of the Ramsey rule and, by construction of the money burning schedule, it is admissible. For $\rho = 0$, the optimal rule

[4] I elaborate on this observation in Online Appendix OA1.

[5] For $\rho < 0$ the penalty would represent a transfer from the agent to the principal. Without a participation constraint, the principal would have an incentive to extract arbitrarily large transfers from the agent.

resembles a Pigouvian tax exactly aligning the incentives of the agent with that of the principal. To see this, note that the optimal rule equates the wedge in the Euler equation of the agent to the marginal discrepancy between the objective of the agent and that of the principal, i.e., $\Delta(\theta, \pi^P(\theta)) = \nu_\pi(\theta, \pi^P(\theta))$, where the Euler equation $\Delta(\theta, \pi) = U^A_\pi(\theta, \pi)$ defines the wedge Δ.

Application to the design of fiscal rules. The principal is the society. The government is the agent in charge of choosing government spending g in response to shocks to the fiscal needs θ of the society.[6] Formally, $\pi \equiv u(g)$ denotes the utility from government spending g, and let the payoff of the principal (before normalization) be $\tilde{U}^P(\theta, \pi) = \theta\pi + W(u^{-1}(\pi) - T)$, where W denotes the continuation value of government deficit $g - T$, and T denotes tax revenues. Although the society and the government agree on the need to respond to shocks to the fiscal needs, the government is present-biased in the sense that it discounts the future at a higher rate than that of the society, $b(\pi) \equiv \beta W(u^{-1}(\pi) - T)$, where $1 - \beta \in [0, 1)$ captures the degree of present bias of the government.[7] Normalizing the payoff of the principal $U^P(\theta, \pi) = \beta\tilde{U}^P(\theta, \pi)$ conveniently implies a bilinear discrepancy between the objectives of the society and the government, $\nu(\theta, \pi) = (1 - \beta)\theta\pi$, and a degree of relative concavity $\kappa = 1$. The parameter ρ is expressed in normalized principal-welfare units, i.e., after the normalization $U^P = \beta\tilde{U}^P$; symmetric money burning corresponds to $\rho = \kappa = 1$ in these units. A fiscal rule is a penalty schedule as a function of government spending, denoted $P(g)$. In the spirit of the revelation principle, a fiscal rule P and an allocation g map into a rule as follows: $\pi(\theta) = u(g(\theta))$ and $\tau(\theta) = P(g(\theta))$ for $\theta \in [\underline{\theta}, \bar{\theta}]$.

The parameter ρ is the cost to society of penalizing the government, relative to the cost borne by the government. One view is that the government responds, in part, to the financial incentives from pay-for-performance contracts (analogous to incentive contracts for central bankers in Walsh (1995)). In that case, τ denotes the reduction in compensation for excessive deficits and the parameter ρ corresponds to the shadow cost of an upper-bound on the expected reduction in compensation that such contracts can provide (see Appendix OA1).[8] In practice, however,

[6] Shocks to government revenues map to shocks to fiscal needs with a CARA utility index (see Section 5.4 of Amador, Werning, and Angeletos (2006)). The non-contractible nature of the shock captures an informational advantage of the governments over society, or heterogeneous information among the members of society as in Piguillem and Schneider (2016).

[7] Amador, Werning, and Angeletos (2006) show that the results apply to a multi-period environment with iid shocks. Halac and Yared (2014) study an infinite horizon environment with persistent shocks.

[8] Note that the current model abstracts from limited enforcement because an upper bound on expected money burning does not bound the severity of off-equilibrium penalties. See Halac and Yared (2022a) for an analysis that allows for the limited enforcement of penalties.

incentive contracts for policy-makers are difficult to implement and not common (Persson and Tabellini (1993)). Instead, the sources of discipline on fiscal policy range from formal enforcement mechanisms such as the Excessive Deficit Procedure of the Stability and Growth Pact to informal mechanisms such as negative coverage by the press. In the context of the U.S., for instance, ρ arguably captures in reduced form the bluntness of sequestration—i.e., automatic across-the-board cuts in government spending (as in Piguillem and Riboni (2024)). Under this interpretation, ρ ranges from 0 to 1 as the sequestration plan shifts from targeted cuts in the government's excess spending to indiscriminate cancelling of government programs. In much of the literature on the design of fiscal rules, the cost of imposing discipline on the fiscal authority is borne symmetrically by the government and the society, in which case $\rho = 1$ (see, for instance, Halac and Yared (2022a)).

2 Solution method

The money-burning constraint limits the applicability of tools commonly used to solve mechanism design problems with transfers. In this section, I extend the applicability of global Lagrangian methods used in Amador, Werning, and Angeletos (2006) and Amador and Bagwell (2013) to obtain necessary and sufficient optimality conditions that completely characterize the solution to delegation problems with money burning.

Although the commonly-used envelope condition remains useful to characterize incentive compatible rules, the standard first-order approach does not apply. The envelope condition conveniently characterizes an incentive-compatible money-burning schedule (Myerson (1981)). The proof is in Online Appendix OA2.1.

Lemma (Incentive compatible allocations)**.** *A rule* (π, τ) *satisfies the incentive compatibility constraint* (2) *if and only if* π *is non-decreasing and*

$$\tau(\theta) = \tau(\underline{\theta}) - U^A(\underline{\theta}, \pi(\underline{\theta})) + U^A(\theta, \pi(\theta)) - \int_{\underline{\theta}}^{\theta} \pi(\tilde{\theta})\, d\tilde{\theta}. \tag{4}$$

This lemma is useful to capture some of the incentive costs of discipline in the objective of the principal. Substituting (4) in (1) transforms the objective of the principal as follows:

$$\int_{\underline{\theta}}^{\bar{\theta}} \left[U^P(\theta, \pi(\theta)) - \rho U^A(\theta, \pi(\theta)) + \rho \pi(\theta) \frac{1 - F(\theta)}{f(\theta)} \right] dF(\theta) - \rho \left(\tau(\underline{\theta}) - U^A(\underline{\theta}, \pi(\underline{\theta})) \right). \tag{5}$$

The assumption on the relative concavity of the objective of the principal to that of the agent, $\rho \leq \kappa$, implies that the integrand in (5) is strictly concave in π for $\rho < \kappa$, and linear in π for the

case of symmetric money burning. Note, however, that the first-order approach—which involves maximizing the transformed objective pointwise—does not apply because (5) does not account for the shadow cost on the money-burning constraint (3).

Unlike the local incentive compatibility constraints, the non-negativity constraints on money burning cannot be easily summarized in the objective function without resorting to Lagrangian methods. To account for the continuum of constraints on money-burning, let the Lagrange multiplier $\Lambda : [\underline{\theta}, \bar{\theta}] \mapsto \mathbb{R}$ be a non-decreasing function. Define the Lagrangian

$$\mathcal{L}(\pi, \underline{\tau}) = \int_{\underline{\theta}}^{\bar{\theta}} \left[U^P(\theta, \pi(\theta)) - \rho\tau(\pi, \underline{\tau})(\theta)\right] dF(\theta) + \rho \int_{\underline{\theta}}^{\bar{\theta}} \tau(\pi, \underline{\tau})(\theta) d\Lambda(\theta) \tag{6}$$

for a non-decreasing policy schedule $\pi : [\underline{\theta}, \bar{\theta}] \mapsto \mathbb{R}$, where $\tau(\pi, \underline{\tau})$ denotes the money burning schedule (4) for incentive compatibility of π given $\tau(\underline{\theta}) = \underline{\tau}$. For future reference, let $\Omega \equiv \{(\pi, \underline{\tau}) | \pi : [\underline{\theta}, \bar{\theta}] \to \mathbb{R} \text{ is non-decreasing and } \underline{\tau} \in \mathbb{R}\}$. The integrals are Lebesgue-Stieltjes integrals.

Given a guess of a rule and a Lagrange multiplier function, global Lagrangian methods provide sufficient optimality conditions to verify the optimality of the rule (Theorem 1 in Amador and Bagwell (2013, 2020)). The condition that the rule maximizes the Lagrangian can be verified using standard tools of optimization based on the derivative of the Lagrangian. Formally, the Gateaux derivative in the direction $(h, h_\tau) \in \Omega$ is

$$\partial\mathcal{L}(\pi, \underline{\tau}, h, h_\tau) \equiv \lim_{\alpha \downarrow 0} \frac{1}{\alpha} \left[\mathcal{L}(\pi + \alpha h, \underline{\tau} + \alpha h_\tau) - \mathcal{L}(\pi, \underline{\tau})\right].$$

The Gateaux derivative exists and, by direct computations, it is[9]

$$\begin{aligned} \partial\mathcal{L}(\pi, \underline{\tau}, h, h_\tau) = &\int_{\underline{\theta}}^{\bar{\theta}} \left[\left(U_\pi^P(\theta, \pi(\theta)) - \rho\Delta(\theta, \pi(\theta)) + \rho\frac{\Lambda(\theta) - F(\theta)}{f(\theta)}\right) h(\theta)\right] dF(\theta) \\ &+ \rho \left[\Delta(\underline{\theta}, \pi(\underline{\theta}))h(\underline{\theta}) - h_\tau\right] \Lambda(\underline{\theta}) + \rho \int_{\underline{\theta}}^{\bar{\theta}} \left[\Delta(\theta, \pi(\theta))h(\theta)\right] d\Lambda(\theta). \end{aligned} \tag{7}$$

The term in parentheses in the first integral is reminiscent of the derivative of the virtual surplus, modified to partially account for the shadow cost of the money-burning constraint. The terms in the second line complete the account of the shadow cost of burning money.

The remaining challenge is to guess the optimal rule and the associated Lagrange multiplier. This is where necessary optimality conditions, which are the main contribution of this section, are useful. There are two requirements for global Lagrangian methods to deliver necessary optimality

[9]The existence of the Gateaux derivative follows from Lemma A.1 in Amador, Werning, and Angeletos (2006). The Lagrange multiplier is normalized so that $\Lambda(\bar{\theta}) = 1$, as justified in the proof of Lemma 1.

conditions. First, the constraint set must have an interior point, which is the case with the possibility of money burning (unlike delegation problems without the possibility to burn money, in which $\tau(\theta) = 0$). Second, the problem must be convex to guarantee that the value function admits a supporting hyperplane, which is the Lagrange multiplier functional. The assumption that $\rho \leq \kappa$ implies that the problem is convex for a broad class of environments, including symmetric money burning. As a result, the global Lagrangian methods guarantee the existence of a Lagrange multiplier functional such that any optimal rule maximizes the Lagrangian and it satisfies a complementary slackness condition.

The difficulty, however, is in obtaining a representation of the Lagrange multiplier functional amenable to using standard tools of optimization. The Riesz representation theorem conveniently guarantees that the Lagrange multiplier functional on the subspace of continuous functions admits an integral representation as in (6). To allow for discontinuities, I use an approximation argument to show that the integral representation is also a supporting hyperplane of the value function on the set of non-decreasing functions. As a result, the next lemma shows that the sufficient optimality conditions are also necessary for the optimality of any rule (including rules with discontinuous policy functions). The proof of the next lemma is in Appendix A.1.

Lemma 1 (Necessary and sufficient conditions for optimality)**.** *A rule (π, τ) is optimal if and only if there exists a non-decreasing function $\Lambda : [\underline{\theta}, \bar{\theta}] \to [0, 1]$ with $\Lambda(\underline{\theta}) = 0$, $\Lambda(\bar{\theta}) = 1$, and Λ right-continuous on $(\underline{\theta}, \bar{\theta})$, such that the following conditions hold:*

(i) For $(h, h_\tau) \in \Omega$, $\partial\mathcal{L}(\pi, \tau(\underline{\theta}), h, h_\tau) \leq 0$.

(ii) In particular, $\partial\mathcal{L}(\pi, \tau(\underline{\theta}), \pi, \tau(\underline{\theta})) = 0$.

(iii) The complementary slackness condition holds: $\int_{\underline{\theta}}^{\bar{\theta}} \tau(\theta) d\Lambda(\theta) = 0$.

(iv) The rule (π, τ) is admissible.

An alternative formulation of the global optimality conditions (i) and (ii) in Lemma 1 greatly simplifies the analysis. The difficulty is that the Gateaux derivative (7) requires evaluating a cross-correlation of the derivative of the virtual surplus and the perturbation function h. An alternative formulation conveniently rewrites the global optimality conditions (i) and (ii) as pointwise optimality conditions. The insight, which I adapt from Amador, Werning, and Angeletos (2006) and Amador and Bagwell (2013) to allow for mechanisms with discontinuities, leverages the monotonicity condition for incentive compatibility. There are two steps. First, integration by parts rewrites the cross-correlation in the Gateaux derivative in terms of the increments of a

perturbation h,[10]

$$-\int_{\underline{\theta}}^{\bar{\theta}} h(\theta)d\delta(\theta) = \int_{\underline{\theta}}^{\bar{\theta}} \delta(\theta)dh(\theta) - \sum_{\theta\in D_{h\delta}} d(\theta) - \delta(\bar{\theta})h(\bar{\theta}) + \delta(\underline{\theta})h(\underline{\theta}), \tag{8}$$

where

$$\begin{aligned}\delta(\hat{\theta}) \equiv \int_{(\hat{\theta},\bar{\theta}]} \left[U_\pi^P(\theta,\pi(\theta)) - \rho\Delta(\theta,\pi(\theta)) + \rho\frac{\Lambda(\theta)-F(\theta)}{f(\theta)}\right] dF(\theta) \\ + \rho\int_{(\hat{\theta},\bar{\theta}]} \Delta(\theta,\pi(\theta))d\Lambda(\theta) + \mathbb{1}(\hat{\theta}=\underline{\theta})\rho\Delta(\underline{\theta},\pi(\underline{\theta}))(\Lambda(\underline{\theta}^+)-\Lambda(\underline{\theta})),\end{aligned}$$

and $D_{h\delta}$ denotes the set of common points of discontinuity of h and δ, and

$$d(\theta) = \left(\delta(\theta) - \frac{\delta(\theta^+)+\delta(\theta^-)}{2}\right)(h(\theta^+)-h(\theta^-)) + \left(\delta(\theta^+)-\delta(\theta^-)\right)\left(h(\theta) - \frac{h(\theta^+)+h(\theta^-)}{2}\right).$$

The set $D_{h\delta}$ is countable because, by definition of Ω, h is non-decreasing.

Substitution of the integral by parts (8) into the Gateaux derivative (7) gives

$$\partial\mathcal{L}(\pi,\underline{\tau},h,h_\tau) = \int_{\underline{\theta}}^{\bar{\theta}} \delta(\theta)dh(\theta) - \sum_{\theta\in D_{h\delta}} d(\theta) + \delta(\underline{\theta})h(\underline{\theta}), \tag{9}$$

where I used that $\Lambda(\underline{\theta}) = 0$ and $\delta(\bar{\theta}) = 0$ to simplify the expression. The second step exploits the monotonicity condition for incentive compatibility to rewrite the global optimality conditions (i) and (ii) in Lemma 1 as a continuum of local optimality conditions. The proof of the next lemma, which uses that a perturbation h can be arbitrarily steep—and even jump—so long as it is non-decreasing, is in Appendix A.2.

Lemma 2 (Pointwise optimality conditions)**.** *Consider a rule (π,τ) and a non-decreasing function $\Lambda : [\underline{\theta},\bar{\theta}] \to [0,1]$ with $\Lambda(\underline{\theta}) = 0$, $\Lambda(\bar{\theta}) = 1$ and Λ right-continuous on $(\underline{\theta},\bar{\theta})$. Conditions (i) and (ii) in Lemma 1 hold if and only if*

(i) $\delta(\theta) \le 0$ for $\theta \in (\underline{\theta},\bar{\theta})$, and $\delta(\underline{\theta}) = \delta(\bar{\theta}) = 0$.

(ii) $\delta(\theta) = 0$ for any $\theta \in [\theta_1,\theta_2)$ where $[\theta_1,\theta_2) \subseteq [\underline{\theta},\bar{\theta}]$ denotes an interval over which π is strictly increasing. At any point θ of discontinuity of π, $\delta(\theta) = 0$ and δ is continuous at θ.

The Lagrange multiplier function has a natural economic interpretation based on a geometric understanding of constrained optimization, where the Lagrange multiplier obtains as a supporting hyperplane to the value function for the principal. The Lagrange multiplier function bounds the

[10] Integration by parts of the Lebesgue-Stieltjes integrals applies because the functions $h(\theta)$ and $\delta(\theta)$ are both functions of bounded variation (Theorem 6.2.2 in Carter and van Brunt (2000)).

change in expected money burning from a change in the money-burning constraint over a range of shocks. To establish this result formally, it is useful to imbed the program of interest in a family of programs parametrized by the tightness of the money-burning constraint. Let $\omega(z)$ denote the value for the principal of a rule that maximizes the principal's objective (1) subject to the incentive compatibility constraint (2) and the more general money-burning constraint $\tau(\theta) \geq z(\theta)$ for $\theta \in [\underline{\theta}, \bar{\theta}]$. The value of an optimal rule to the principal is $\omega(0)$. Because the Lagrange multiplier functional obtains as a supporting hyperplane to ω, it provides a lower bound for the cost to the principal of a tightening of the money-burning constraint

$$\omega(0) - \omega(z) \geq \rho \int_{\underline{\theta}}^{\bar{\theta}} z(\theta) d\Lambda(\theta).$$

More generally, if ω is Gateaux differentiable at 0 in the direction z, then

$$\partial\omega(0, z) = -\rho \int_{\underline{\theta}}^{\bar{\theta}} z(\theta) d\Lambda(\theta).$$

In particular, if the value function ω is Gateaux differentiable at 0 in the direction of tightening the money-burning constraint for $\theta \leq \theta_1$, then $z(\theta) = 1$ for $\theta \leq \theta_1$ and $z(\theta) = 0$ otherwise, and $\partial\omega(0, z) = -\rho\Lambda(\theta_1)$. For instance, a uniform tightening of the money-burning constraint by a constant $\bar{z}$ changes the value to the principal by $\omega(\bar{z}) - \omega(0) = -\rho\bar{z}$. This is because a uniform tightening of the money-burning constraint only shifts the money-burning schedule uniformly.

3 Necessary optimality conditions

The main results in this section apply Lemma 1 and Lemma 2 to obtain easy-to-check necessary conditions for the optimality of marginal deterrence, bunching, and bounded discretion.

3.1 Marginal deterrence

Marginal deterrence can strike a fine balance between discipline and flexibility at the cost of burning money. The next proposition shows that, marginal deterrence can be part of an optimal rule only if it satisfies a simple formula that determines the optimal balance between discipline and flexibility at the margin. In turn, to be compatible with incentives, the formula for marginal deterrence must implement a non-decreasing policy schedule, which implies a condition on the thickness of the tail of the distribution of shocks. For future reference, say that condition $c_m(\theta_1, \theta_2, \pi, \lambda)$ holds if, for $\theta \in (\theta_1, \theta_2)$,

$$(1 - \rho)\Delta(\theta, \pi(\theta)) = \nu_\pi(\theta, \pi(\theta)) - \rho\frac{\lambda - F(\theta)}{f(\theta)}, \tag{10}$$

and

$$\rho \frac{d}{d\theta} \frac{\lambda - F(\theta)}{f(\theta)} \geq \rho - U^P_{\theta\pi}(\theta, \pi(\theta)). \tag{11}$$

Proposition 1 (Marginal deterrence)**.** *Let (π, τ) be an optimal rule. If $\tau(\theta) > 0$ for $\theta \in (\theta_1, \theta_2)$ and π is strictly increasing over (θ_1, θ_2), then there exists $\lambda \in [0, 1]$ such that condition $c_m(\theta_1, \theta_2, \pi, \lambda)$ holds. If $\tau(\theta) > 0$ for $\theta \in [\underline{\theta}, \theta_2)$, then $\lambda = 0$. If $\tau(\theta) > 0$ for $\theta \in (\theta_1, \bar{\theta}]$, then $\lambda = 1$.*

The proof is in Appendix A.3. Condition (10) characterizes a family—parametrized by $\lambda \in [0, 1]$—of optimal wedges in the Euler equation of the agent induced by marginal deterrence. Under the assumption on the relative concavity of the objective of the principal and the agent $\rho < \kappa$, and by continuity of f, F, U^A_π, and U^P_π, equation (10) implicitly defines a unique and continuous policy schedule parametrized by λ. The constant λ is the value taken over (θ_1, θ_2) by the non-decreasing Lagrange multiplier Λ—which the complementary slackness condition holds fixed wherever the rule burns money—so that λ captures the shadow cost of the money-burning constraint over that interval. The Lagrange multiplier may take an interior value, i.e., $\lambda \in (0, 1)$, only if the rule burns money in the interior of the range of shocks. In that case, marginal deterrence is characterized by three parameters, $\theta_1, \theta_2 \in (\underline{\theta}, \bar{\theta})$, and $\lambda \in (0, 1)$, such that the conditions (10), $\tau(\theta_1) = 0$, and $\tau(\theta_1) = \tau(\theta_2)$ hold, where τ satisfies the envelope condition (4).

For environments with symmetric money burning, however, marginal deterrence can be part of an optimal rule only under a knife-edge condition on the bias and the distribution of shocks. This is because the constant relative concavity of U^A and U^P implies that the objective of the principal takes the form

$$U^P(\theta, \pi) = \kappa[U^A(\theta, \pi) + B(\theta) + C(\theta)\pi]$$

and, for $\rho = \kappa$, equation (10) reduces to $C(\theta) = -\frac{\lambda - F(\theta)}{f(\theta)}$ for some $\lambda \in [0, 1]$. Thus bang-bang incentives are a generic property of optimal rules in a general principal-agent model with symmetric money burning (see Proposition 1 in Halac and Yared (2022a) where this result first appeared for the design of a fiscal rule in a model where $C(\theta) = -(1 - \beta)\theta$.)

Application to the design of fiscal rules. If an optimal fiscal rule deters spending at the margin (i.e., π is strictly increasing and $\pi \neq \pi_f$ over an interval), then Proposition 1 implies that the optimal marginal penalty induces the following wedge in the government's Euler equation

$$\Delta(\theta, \pi(\theta)) = \frac{1}{1 - \rho} \left((1 - \beta)\theta - \rho \frac{\lambda - F(\theta)}{f(\theta)} \right)$$

for some $\lambda \in [0,1]$. In some cases, it is possible to determine λ without completely solving for the optimal rule. In particular, if the present-bias is sufficiently large relative to the inverse of the elasticity of the distribution of shocks in the sense that $1-\beta \geq \rho\frac{1-F(\theta)}{\theta f(\theta)}$ for $\theta \in (\theta_1, \bar{\theta}]$, then the marginal penalty associated with marginal deterrence is necessarily positive, and $\lambda = 1$ because $\tau(\theta) > 0$ for $\theta \in (\theta_1, \bar{\theta}]$. Note that this condition is sufficient but not necessary for $\lambda = 1$.

The formula for the wedge is the product of a scaling factor and a term capturing the key trade-off between discipline and flexibility in marginal deterrence. The scaling factor $1/(1-\rho)$ measures society's aversion to penalizing the government. It is the inverse of the degree of asymmetry in the cost of enforcement and ranges between 1 and infinity. It takes its smallest value for the largest degree of asymmetry. Indeed, for $\rho = 0$, society does not bear any cost associated with penalizing the government. In contrast, for a symmetric cost of enforcement (i.e., $\rho = \kappa = 1$), society is infinitely averse to burning money, and only the sign of the term in parentheses matters.[11]

The term in parentheses captures the key trade-off between discipline and flexibility in marginal deterrence. The term $(1-\beta)\theta$ captures the need to discipline a deficit-biased government. The term $\rho\frac{1-F(\theta)}{f(\theta)}$ captures the *incentive cost* of marginal deterrence. To be compatible with incentives, marginal penalties are cumulative in the sense that a marginal penalty on spending $g(\theta)$ is borne with weight ρ and probability $1-F(\theta)$. It also relaxes the money-burning constraint, which has value $\rho(1-\lambda)$ to the principal.

The monotonicity condition (11) with $\lambda = 1$ is a lower bound on the thickness of the right-tail of the distribution of shocks, $\rho\frac{d}{d\theta}\frac{1-F(\theta)}{f(\theta)} \geq \rho - \beta$. Intuitively, a thick right-tail of the distribution of shocks calls for a fine balance between discretion and discipline that only marginal deterrence can achieve. The threshold for the thickness of the right-tail of the distribution of shocks depends on the degree of asymmetry in the cost of enforcement and the degree of present bias. For $\rho = \beta$, it suffices to check whether the inverse hazard rate is increasing, which is the case if and only if $1-F$ is log-convex. For instance, for $\rho = \beta$, the monotonicity condition is satisfied for Pareto-distributed shocks to the fiscal needs whose inverse hazard rate is increasing, but fails for exponentially distributed shocks to fiscal needs on a truncated support because the inverse hazard rate is strictly decreasing.

[11] This intuition echoes the bang-bang result of Halac and Yared (2022a): generically, an optimal rule enforced by symmetric penalties necessarily relies on extreme—bang-bang—penalties. The wedge generalizes the function $Q(\theta) = -(1-\beta)\theta f(\theta) + 1 - F(\theta)$ studied in Halac and Yared (2022a) to study environments with asymmetric penalties.

3.2 Bunching

A rule can impose discipline by limiting the set of implementable policies. A cap, a floor, or a hole in the range of permissible actions induces the bunching of policies over a range of shocks. More generally, bunching can occur either because of a kink or a notch in the penalty schedule.

A difficulty in characterizing bunching intervals in an optimal rule is that the possible patterns of bunching are varied and the policy schedule can be discontinuous (see, for instance, the escalator-and-stairs class of mechanisms studied in Amador, Bagwell, and Carpizo (2026)). In particular, if the agent's incentives regarding the prescribed policy change sign over the bunching interval, the shadow cost of money burning may jump. The main contribution of this section is a set of necessary optimality conditions for bunching in any optimal rule (i.e., irrespective of whether the rule burns money or not and allowing for discontinuous rules). For future reference, say that condition $c_b(\theta_1, \theta_2, \pi^*, \underline{\lambda}, \bar{\lambda})$ holds if, for the step function $\lambda : [\theta_1, \theta_2] \to \{\underline{\lambda}, \bar{\lambda}\}$, $\lambda(\theta_1) = \underline{\lambda}$, $\lambda(\theta_2) = \bar{\lambda}$, and for $\theta \in (\theta_1, \theta_2)$

$$\lambda(\theta) = \begin{cases} \bar{\lambda} & \text{if } \pi^* \leq \pi_f(\theta), \\ \underline{\lambda} & \text{if } \pi^* > \pi_f(\theta), \end{cases}$$

the following inequality

$$(1-\rho)\int_{\theta}^{\theta_2} \Delta(\hat{\theta}, \pi^*) dF(\hat{\theta}) \leq \int_{\theta}^{\theta_2} \left[\nu_\pi(\hat{\theta}, \pi^*) - \rho \frac{\bar{\lambda} - F(\hat{\theta})}{f(\hat{\theta})} \right] dF(\hat{\theta}) - \rho \Delta(\theta, \pi^*)(\bar{\lambda} - \lambda(\theta)) \tag{12}$$

holds for $\theta \in [\theta_1, \theta_2]$; and inequality (12) holds as an equality for $\theta = \theta_1$.

Proposition 2 (Bunching)**.** *Let (π, τ) be an optimal rule. If there exists $\pi^* \in (\underline{\pi}, \bar{\pi})$ and $\theta_1, \theta_2 \in [\underline{\theta}, \bar{\theta}]$, $\theta_1 < \theta_2$, such that $\pi(\theta) = \pi^*$ for $\theta \in (\theta_1, \theta_2)$ and $\pi(\theta) \neq \pi^*$ for $\theta \notin [\theta_1, \theta_2]$, then there exist $\underline{\lambda}, \bar{\lambda} \in [0, 1]$, with $\underline{\lambda} \leq \bar{\lambda}$ such that condition $c_b(\theta_1, \theta_2, \pi^*, \underline{\lambda}, \bar{\lambda})$ holds. If $\tau(\theta) > 0$ for $\theta \in (\theta_1, \theta_2)$, then $\underline{\lambda} = \bar{\lambda}$. If $\theta_1 = \underline{\theta}$, then $\underline{\lambda} = 0$. If $\theta_2 = \bar{\theta}$, then $\bar{\lambda} = 1$.*

The proof is in Appendix A.4.

Despite the single optimality condition c_b, there are two distinct economic motives for bunching. The Lagrange multiplier function over the bunching interval distinguishes bunching to iron a non-monotonic policy plan to preserve incentive compatibility, from bunching due to an exemption from marginal deterrence. Consider the standard case of ironing. Suppose that the multiplier does not jump, so that $\underline{\lambda} = \bar{\lambda} = \lambda$. Condition c_b has two parts: an equality and a continuum of inequalities. First, evaluating inequality (12) at $\theta = \theta_1$, where it holds with equality, shows that the wedge over the bunching interval equals the expected wedge from marginal deterrence. That is, bunching irons the wedge that marginal deterrence would otherwise induce.

At any θ in the bunching interval, if inequality (12) did not hold, the principal would benefit from offering an alternative rule with a kink in the penalty schedule at π^* that preserves the bunching between θ_1 and θ while marginally raising the policy schedule between θ and θ_2 (see Figure 1 for an example). Consider next the case in which the multiplier jumps at $\theta \in (\theta_1, \theta_2)$. By definition of the step function λ, the wedge changes sign, that is, where $\pi^* = \pi_f(\theta)$ and $\Delta(\theta, \pi^*) = 0$. Bunching continues to iron the wedge of marginal deterrence, now at a higher multiplier above the sign change than below it.

Consider next the case of an exemption from marginal deterrence. In that case, the multiplier instead jumps where the wedge differs from zero. Because the step function ties the multiplier to the sign of the wedge in the interior of the range of shocks, such a jump can occur only at $\underline{\theta}$ or $\bar{\theta}$, where the boundary conditions $\underline{\lambda} = 0$ and $\bar{\lambda} = 1$ fix the multiplier irrespective of the sign of the wedge. The resulting kink in the penalty schedule reflects an exemption from marginal deterrence, as the application to the design of fiscal rules illustrates (Corollary 2 and Figure 2).

Application to the design of fiscal rules. Because ν is linear in π, the continuum of inequalities (12) is easy to verify directly from three fundamentals of the economy, namely, the degree of present bias $1 - \beta$, the degree of asymmetry in the cost of enforcement $1 - \rho$, and the conditional tail expectation of the distribution of shocks.

As the next corollary shows, the equality condition in c_b sets the level of bunching (were the inequality slack at $\theta = \theta_1$, a lower π^* would raise welfare and vice versa), and the continuum of inequality conditions in c_b relates to the thickness of the tail of the distribution of shocks.

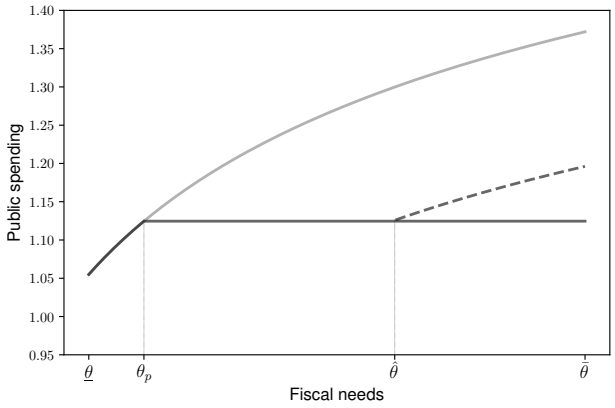

Figure 1: On the continuum of inequalities in the definition of the thresholds θ_p.

Notes: The grey line depicts the flexible spending schedule. The black line depicts an allocation π implemented by a cap at $\pi(\theta_p)$. The dashed line depicts an alternative to the cap. To implement the alternative, the fiscal rule resorts to marginal deterrence after a kink at $\pi(\theta_p)$.

Corollary 1 (Necessary optimality conditions for a cap)**.** *Let (π, τ) be an optimal fiscal rule.*

If there is $\theta_p \in (\underline{\theta}, \bar{\theta})$ such that $\pi(\theta) = \pi(\theta_p)$ for $\theta \geq \theta_p$ and $\pi(\theta_p) \leq \pi_f(\theta_p)$, then the threshold θ_p solves

$$(1-\rho)\Delta(\theta_p, \pi(\theta_p)) = \theta_p - \beta E[\theta|\theta \geq \theta_p],$$

and

$$\beta E\big[\hat{\theta}\big|\hat{\theta} \geq \theta\big] - \rho\theta \leq \beta E\big[\hat{\theta}\big|\hat{\theta} \geq \theta_p\big] - \rho\theta_p$$

for every $\theta \in [\theta_p, \bar{\theta}]$.

If the fiscal rule leaves no discretion (i.e., π is constant), then $\pi(\theta) = \pi^{ER}$ and, for $\theta \in [\underline{\theta}, \bar{\theta}]$

$$\beta E\big[\hat{\theta}|\hat{\theta} \geq \theta\big] - \rho\theta \leq \beta E\big[\hat{\theta}\big](1-\rho).$$

A well-designed cap equates the average distortion at the top on the right-hand side to the benefit of not imposing the marginal penalty at $\pi(\theta_p)$ with probability $1 - F(\theta_p)$ on the left-hand side. For a rule that does not burn money, the left-hand side is null because the wedge is null. Then the cap is such that there is no distortion at the top on average. If, instead, the cap comes on top of marginal deterrence, then the left-hand side is positive.

An optimal fiscal rule may also induce the bunching of the allocation due to a kink in the penalty schedule resulting from an exemption from marginal deterrence below a threshold.

Corollary 2 (Necessary optimality conditions for an exemption from penalties below a threshold)**.** *If there is $\theta_x \in (\underline{\theta}, \bar{\theta})$ such that $\tau(\theta) > 0$ for $\theta > \theta_x$ and $\pi(\theta) = \pi(\theta_x) \leq \pi_f(\underline{\theta})$ for $\theta \leq \theta_x$, then*

$$(1-\rho)\Delta(\theta, \pi(\theta_x)) \geq \theta - \beta E[\hat{\theta}|\hat{\theta} \leq \theta] - \rho\frac{\Delta(\theta, \pi(\theta_x))}{F(\theta)}$$

for every $\theta \in (\underline{\theta}, \theta_x]$, and the inequality holds as an equality for $\theta = \theta_x$.

To gain insights into the optimality of granting an exemption from marginal deterrence below a threshold, decompose the penalty schedule into two building blocks: the intercept and the marginal penalty schedule. While the marginal penalty schedule determines the extent of discipline, the intercept determines the overall burden of penalties. The non-negativity constraint on penalties forces the intercept to be non-negative; hence, the overall burden of penalties can only be shifted up, which is not desirable. The only option to lower the overall burden of penalties is to alter the marginal penalty schedule. Because lowering the marginal penalty on an action π lowers the level of penalties on all actions above π, altering the bottom of the marginal penalty schedule is the most effective substitute for a negative intercept. Unlike a negative intercept, however, the truncation entails a loss of discipline because it grants an *exemption* from marginal

deterrence below a threshold. The exemption causes a jump in the marginal penalty schedule, and the resulting kink induces the bunching of government spending at the exemption threshold (see Figure 2).

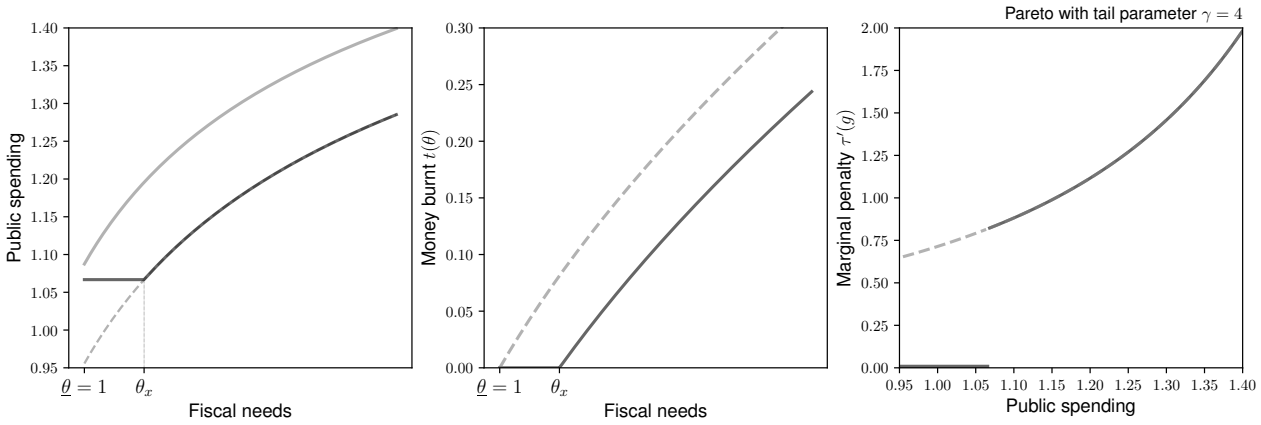

Figure 2: Exemption from penalties below $\pi(\theta_x)$.

Notes: The grey line in the left panel depicts the discretionary allocation. The dashed lines and the black lines depict the allocation (left panel), the money burnt (middle panel), and the marginal penalties (right panel) with and without the exemption from penalties for public spending below $g(\theta_x)$. While the exemption forgoes discipline below $g(\theta_x)$ (left panel) by truncating the marginal penalty schedule (right panel), it lowers the level of penalty (middle panel) while keeping the same level of discipline above $g(\theta_x)$ (left and right panels).

3.3 Discretion

Alternatively, a rule may grant discretion over a range of policies. The next proposition shows that an optimal rule grants discretion only under a condition on the distribution of shocks and the discrepancy between the payoff of the principal and that of the agent.[12] For future reference, say that condition $c_d(\theta_1, \theta_2)$ holds if

$$\nu_\pi(\theta, \pi_f(\theta))f(\theta) + \rho F(\theta)$$

is non-decreasing over (θ_1, θ_2).

Proposition 3 (Discretion)**.** *Let (π, τ) be an optimal rule. If $\pi(\theta) = \pi_f(\theta)$ for $\theta \in (\theta_1, \theta_2)$, then condition $c_d(\theta_1, \theta_2)$ holds. If $\theta_1 = \underline{\theta}$, then $\nu_\pi(\underline{\theta}, \pi_f(\underline{\theta})) \geq 0$. If $\theta_2 = \bar{\theta}$, then $\nu_\pi(\bar{\theta}, \pi_f(\bar{\theta})) \leq 0$. If $\rho = 0$ and $\pi(\theta) = \pi_f(\theta)$ for $\theta \in (\theta_1, \theta_2)$, then $\nu_\pi(\theta, \pi_f(\theta)) = 0$ for $\theta \in (\theta_1, \theta_2)$.*

[12]Proposition 3 extends previous results on the necessity of condition c_d for the optimality of interval delegation to any optimal rule, including discontinuous rules and rules that burn money. Condition c_d corresponds to condition (c1) in Proposition 2 (b) and Section 4.3 in Amador and Bagwell (2013).

The proof, which follows from the pointwise optimality condition (ii) in Lemma 2, is in Appendix A.5. Condition (ii) in Lemma 2 imposes that the Lagrange multiplier function satisfies $\rho\Lambda(\theta) = \rho F(\theta) + \nu_\pi(\theta, \pi_f(\theta))f(\theta)$ over any interval over which $\pi(\theta) = \pi_f(\theta)$. For $\rho > 0$, this identifies $\Lambda(\theta) = F(\theta) + \frac{1}{\rho}\nu_\pi(\theta, \pi_f(\theta))f(\theta)$, and the requirement that Λ be non-decreasing is condition c_d. For $\rho = 0$, the equation instead collapses to $\nu_\pi(\theta, \pi_f(\theta)) = 0$, the final clause of the proposition.

The intuition for this condition relates to the cumulative shadow cost of the money-burning constraint. If the condition is not satisfied at $\theta \in (\theta_1, \theta_2)$, then the money burning-constraint is not binding locally at θ because a steepening perturbation of the allocation in a neighborhood of $\pi_f(\theta)$ improves welfare. For the case of symmetric money burning, because optimal incentives are necessarily bang-bang, the steepening perturbation "drills a hole" by removing an interval in the range of implementable policies to avoid burning money. As shown in the proof of Lemma 1, a discontinuous allocation can be approximated arbitrarily closely by an allocation that burns marginally more money. For asymmetric money burning, i.e., $\rho < 1$, the necessary condition is more stringent because steepening perturbations are less costly to the principal.

The necessary optimality conditions in Proposition 3 must also imply that flattening perturbations of the policy schedule weakly reduce welfare. As the next claim shows, the optimality condition c_d implies that flattening perturbations reduce welfare because the incentive cost is large relative to the bias between the principal and the agent.

Corollary 3 (The optimality of granting discretion implies a bound on the bias)**.** *Let* (π, τ) *denote an optimal rule. If* $\pi(\theta) = \pi_f(\theta)$ *for* $\theta \in (\theta_l, \theta_p)$*, then*

$$-\rho\frac{F(\theta)}{f(\theta)} \leq \nu_\pi(\theta, \pi_f(\theta)) \leq \rho\frac{1 - F(\theta)}{f(\theta)} \quad \textit{for } \theta \in (\theta_l, \theta_p).$$

The proof is in Online Appendix OA2.3. The upshot of Corollary 3 is a simple test that can reject the optimality of interval delegation.

Application to the design of fiscal rules. In the context of the fiscal policy model with $\rho > 0$, the necessary condition c_d for the optimality of granting discretion over $[\pi_f(\theta_1), \pi_f(\theta_2)]$ simplifies to a lower bound on the elasticity of the density of the distribution of shocks,

$$\theta\frac{f'(\theta)}{f(\theta)} \geq -\frac{1 - \beta + \rho}{1 - \beta}$$

for $\theta \in [\theta_1, \theta_2]$. The condition generalizes a condition found in much of the literature on symmetric money burning to environments with asymmetric money burning (cf. Assumption A in Amador,

Werning, and Angeletos (2006), and the condition that $Q'(\theta) < 0$ in Assumption 1 in Halac and Yared (2022a)).

Corollary 3 shows that an optimal rule grants discretion below a threshold, say $\pi_f(\theta_p)$, only if the degree of present bias is bounded above by the inverse of the elasticity of the tail of the distribution of shocks, $1 - \beta \leq \rho \frac{1-F(\theta)}{\theta f(\theta)}$ for $\theta \leq \theta_p$.

4 Optimal rules

In this section, I solve for globally optimal rules in two steps. First, I obtain sufficient optimality conditions by piecing together necessary optimality conditions. I consider two broad classes of rules that are commonly used in practice. The first class of rules provides bounded discretion by delegating an interval. The second class of rules imposes a graduated schedule of penalties above a threshold. While delegating an interval imposes discipline without burning money, a graduated schedule of penalties strikes a fine balance between discipline and discretion at the cost of burning money. Second, I apply the theory to solve for globally optimal fiscal rules and study comparative statics with respect to the distribution of shocks and the severity of the bias.

The main takeaway from this section is that, to build a globally optimal rule, it suffices to find a non-decreasing policy plan that is locally optimal within each piece of the partition, and such that the associated Lagrange multiplier is non-decreasing.

4.1 Interval delegation

A main result of the delegation literature is that, under some conditions on the distribution of shocks, interval delegation is optimal. A rule (π, τ) *delegates an interval* parametrized by $\theta_l, \theta_p \in [\underline{\theta}, \bar{\theta}]$ with $\theta_l < \theta_p$,

$$\pi(\theta) = \min(\pi_f(\theta_p), \max(\pi_f(\theta_l), \pi_f(\theta))),$$

and $\tau(\theta) = 0$ for $\theta \in [\underline{\theta}, \bar{\theta}]$. This definition does not cover the limit case of a degenerate interval, in which case an optimal rule necessarily implements the expected Ramsey allocation. A rule (π, τ) is a *narrow mandate* if $\pi(\theta) = \pi^{ER}$ and $\tau(\theta) = 0$ for $\theta \in [\underline{\theta}, \bar{\theta}]$.

As the next corollary shows, sufficient optimality conditions for interval delegation obtain by piecing together necessary optimality conditions with a non-decreasing Lagrange multiplier function. The Lagrange multiplier function that pieces together the necessary optimality conditions is pinned down by the necessary optimality conditions. Using that $\Lambda(\underline{\theta}) = 0$ and $\Lambda(\bar{\theta}) = 1$,

the necessary optimality condition for bunching c_b implies that the Lagrange multiplier function is constant at its minimal value over the bunching interval $[\underline{\theta}, \theta_l)$, and constant at its maximal value over the bunching interval $(\theta_p, \bar{\theta}]$. The proof of the necessary optimality condition c_d gives the Lagrange multiplier function over $[\theta_l, \theta_p]$. Lastly, the necessary optimality conditions jointly imply that the Lagrange multiplier is non-decreasing at the boundary points θ_l and θ_p (as shown in the proof in Online Appendix OA2.4).

Corollary 4 (Optimality of interval delegation)**.** *1. A rule (π, τ) delegating an interval parametrized by $\theta_l, \theta_p \in [\underline{\theta}, \bar{\theta}]$, $\theta_l < \theta_p$ is optimal if and only if i) condition $c_b(\underline{\theta}, \theta_l, \pi_f(\theta_l), \underline{\lambda} = 0, \bar{\lambda} = 0)$ holds, or $\theta_l = \underline{\theta}$ and $\nu_\pi(\underline{\theta}, \pi(\underline{\theta})) \geq 0$; ii) condition $c_d(\theta_l, \theta_p)$ holds; and iii) condition $c_b(\theta_p, \bar{\theta}, \pi_f(\theta_p), \underline{\lambda} = 1, \bar{\lambda} = 1)$ holds, or $\theta_p = \bar{\theta}$ and $\nu_\pi(\bar{\theta}, \pi(\bar{\theta})) \leq 0$.*

2. A rule (π, τ) delegating a narrow mandate is optimal if and only if condition $c_b(\underline{\theta}, \bar{\theta}, \pi^{ER}, \underline{\lambda} = 0, \bar{\lambda} = 1)$ holds.

Application to the design of fiscal rules. In this application, the optimality conditions do not depend on the details of the utility function for the agent because the discrepancy between the payoff of the government and society is bilinear in the shock and the policy, i.e., $\nu(\theta, \pi) = (1 - \beta)\theta\pi$. The distribution of shocks, the degree of present bias β, and the cost of burning money ρ alone determine the optimal fiscal rule. For simplicity, suppose that $\beta = \rho$ so that the optimality conditions can be easily checked for a log-concave tail $1 - F$. For $\rho > \beta$, the optimality conditions for bunching and discretion are relaxed, and vice versa for $\rho < \beta$. Suppose that the tail of the distribution of shocks is log-concave. Specifically, let fiscal needs be distributed according to the exponential distribution with scale parameter $\lambda_d > 0$, truncated and shifted on the support $[1, \bar{\theta}]$.

We can rule out marginal deterrence as suboptimal for a log-concave distribution and $\beta = \rho$. For marginal deterrence to be part of an optimal rule, the monotonicity condition (11) would need to hold for some $\lambda \in [0, 1]$. Note however, that

$$\rho \frac{d}{d\theta} \frac{\lambda - F(\theta)}{f(\theta)} \leq \rho \frac{d}{d\theta} \frac{1 - F(\theta)}{f(\theta)} < 0 = \rho - \beta,$$

where the first inequality follows from $f' < 0$, $\lambda \in [0, 1]$, and the second inequality follows because the inverse hazard rate of the truncated exponential distribution is strictly decreasing.

Second, a complete characterization of the optimal rule obtains based on Corollary 4. Formally, there is a threshold degree of present bias above which the optimal fiscal rule is a narrow mandate, and below which the optimal fiscal rule grants some discretion below a cap on spending. For truncated-exponentially distributed shocks, a narrow mandate is optimal if the degree of present

bias is sufficiently high that

$$\frac{1-\beta}{\beta} \geq \frac{1}{\lambda_d} - \frac{\bar{\theta}-1}{e^{\lambda_d(\bar{\theta}-1)}-1}. \tag{13}$$

This follows from Corollary 4.2. Recall that, for a bilinear bias function ν, condition $c_b(\underline{\theta}, \bar{\theta}, \pi^{ER}, \underline{\lambda} = 0, \bar{\lambda} = 1)$, reduces to the inequality $\beta E\big[\theta|\theta \geq \hat{\theta}\big] - \rho\hat{\theta} \leq \beta E\big[\theta\big](1-\rho)$ holding for $\hat{\theta} \in [\underline{\theta}, \bar{\theta}]$. For $\beta = \rho$, the decreasing mean residual life of the truncated exponential distribution implies that if the inequality holds for $\theta = \underline{\theta}$, then it holds for $\theta \geq \underline{\theta}$. The inequality evaluated at $\theta = \underline{\theta}$ gives the previous inequality condition on the degree of present bias.

If the reverse inequality holds instead,

$$\frac{1-\beta}{\beta} < \frac{1}{\lambda_d} - \frac{\bar{\theta}-1}{e^{\lambda_d(\bar{\theta}-1)}-1},$$

then the optimal fiscal rule grants some discretion below a cap. The threshold for the cap solves

$$\theta_p = \frac{\beta}{1-\beta}\left(\frac{1}{\lambda_d} - \frac{\bar{\theta}-\theta_p}{e^{\lambda_d(\bar{\theta}-\theta_p)}-1}\right),$$

and the condition on the degree of present bias implies that the cap leaves some discretion, i.e., $\theta_p \in (1, \bar{\theta}]$. Condition (iii) in Corollary 4.1 is satisfied because the mean residual life of a truncated exponential distribution is decreasing. The floor is not binding—i.e., $\theta_l = \underline{\theta}$—and condition (i) in Corollary 4.1 holds. If, in addition, condition (ii) for the optimality of granting discretion is satisfied, then the optimal fiscal rule grants discretion below the cap. Condition (ii) in Corollary 4.1 is an upper bound on the bias relative to the elasticity of the density. Because the elasticity of the density of a (truncated) exponential distribution is decreasing, it suffices that $(1-\beta)\theta_p \leq \frac{1}{\lambda_d}$ for the three sufficient optimality conditions in Corollary 4.1 to be satisfied.

The top two panels in Figure 3 illustrate this characterization for log utility, scale parameter $\lambda_d = 3$, and $\bar{\theta} = 4$. In the right panel, $\beta = 0.7$ and a narrow mandate is optimal because $\frac{1-\beta}{\beta} > \frac{1}{\lambda_d} - \frac{\bar{\theta}-1}{e^{\lambda_d(\bar{\theta}-1)}-1} \approx 0.333$. In the left panel, $\beta = 0.8$ and a cap on spending that leaves some discretion is an optimal fiscal rule because $\frac{1-\beta}{\beta} < \frac{1}{\lambda_d} - \frac{\bar{\theta}-1}{e^{\lambda_d(\bar{\theta}-1)}-1} \approx 0.333$. The condition for the optimality of discretion below the cap holds, $0.27 \approx (1-\beta)\theta_p \leq \frac{1}{\lambda_d} \approx 0.33$ and the threshold $\theta_p \approx 1.33$.

4.2 Graduated schedule of penalties

Unlike interval delegation, which disciplines the agent by effectively limiting the choice set, the rules studied in this section strike a fine balance between discipline and flexibility at the cost of burning money. This section focuses on environments with asymmetric money burning (i.e., $\rho < \kappa$). Recall that, for symmetric money burning, Proposition 1 implies that marginal deterrence

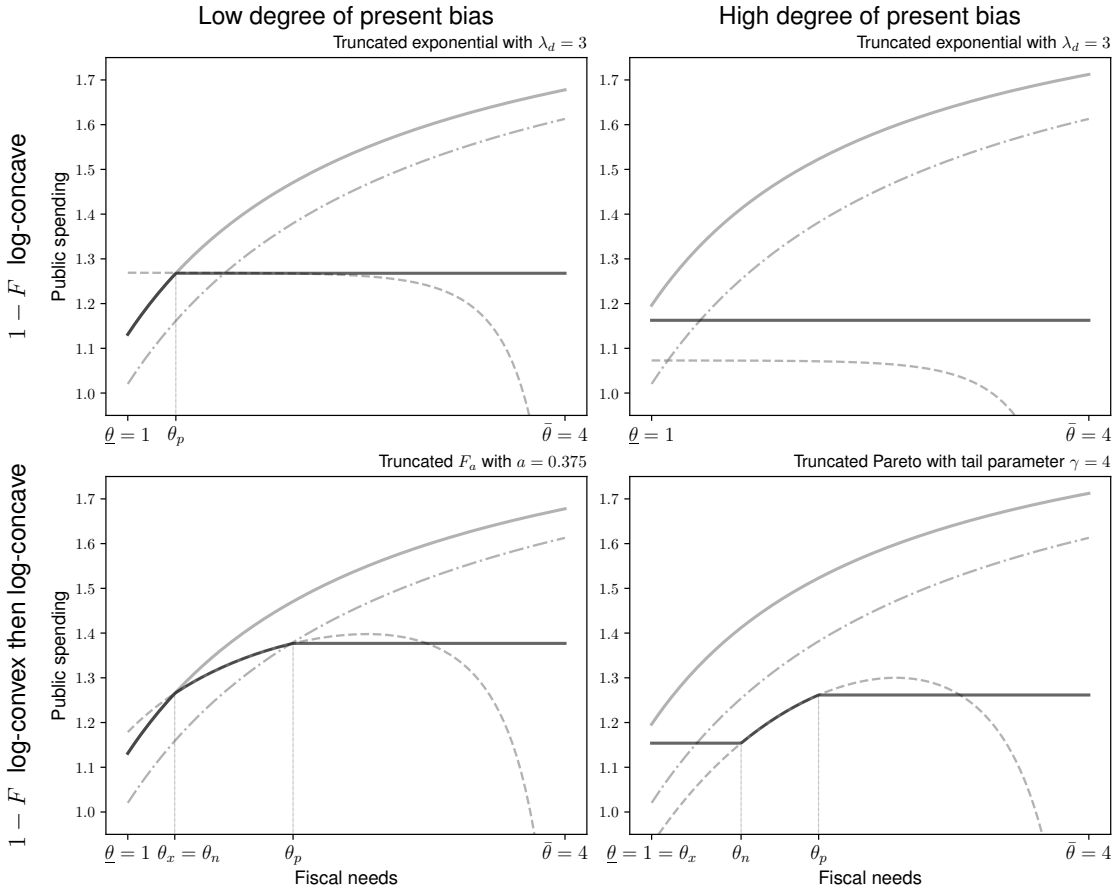

Figure 3: Optimal fiscal rules under a low and a high degree of present bias.

Notes: Each panel plots public spending as a function of fiscal needs θ. The solid grey line depicts the government's preferred policy in the absence of a fiscal rule whereas the dash-dotted line depicts society's preferred policy. The dashed line depicts the allocation that satisfies the wedge (10) in the necessary optimality condition for marginal deterrence with $\lambda = 1$, repeated here for convenience $\Delta(\theta, \pi(\theta)) = \frac{1}{1-\rho}\left((1-\beta)\theta - \rho\frac{1-F(\theta)}{f(\theta)}\right)$. The solid black line is the policy plan implemented by the optimal fiscal rule. The present bias parameter is $\beta = 0.8$ for the panels in the left column, and $\beta = 0.7$ for the panels in the right column, with $\beta = \rho$ throughout and log utility $U(g) = \ln(g)$. The distribution of shocks is at the top-right of each panel.

can only be optimal under a knife-edge condition on the bias and the distribution of shocks. That is, generically the principal disciplines the agent by restricting the choice set, possibly with discontinuous policy schedules (as in the escalator-and-stairs rules studied in Amador, Bagwell, and Carpizo (2026)). For rules with asymmetric money burning, by contrast, the strict relative concavity condition $\rho < \kappa$ implies that optimal rules are continuous.

Lemma 3 (Asymmetric money burning and the continuity of optimal rules)**.** *Suppose $\rho < \kappa$. Then an optimal rule (π, τ) implements a continuous policy schedule π.*

The proof is in Appendix OA2.5. Lemma 3 implies that, for $\rho < \kappa$, restricting attention to continuous policy schedules is without loss of generality.

A rule (π, τ) imposes a *graduated schedule of penalties* if π and τ are absolutely continuous, non-decreasing, there exists $\theta_n \in (\underline{\theta}, \bar{\theta})$ such that $\tau(\theta) = 0$ for $\theta < \theta_n$ and $\tau(\theta) > 0$ for $\theta > \theta_n$, and there exists $\theta_x \in [\underline{\theta}, \theta_n]$ such that $\pi(\theta) = \pi_f(\theta)$ for $\theta \in [\underline{\theta}, \theta_x)$ and $\pi(\theta) = \pi(\theta_n)$ for $\theta \in [\theta_x, \theta_n]$. Examples of policy schedules implemented by a graduated schedule of non-decreasing penalties are depicted in the bottom panels of Figure 3.

Sufficient optimality conditions obtain by piecing together necessary optimality conditions with a non-decreasing Lagrange multiplier. Let (π, τ) denote a rule that imposes a graduated schedule of penalties above a threshold parametrized by θ_x and θ_n. There are four necessary optimality conditions. For shock realizations above θ_n, there can be either marginal deterrence or bunching. Proposition 1 indicates that the following conditions must hold over any interval of shocks over which marginal deterrence leaves some flexibility to respond to shocks:

(c1) let $(\theta_1, \theta_2) \subseteq (\theta_n, \bar{\theta}]$ denote an interval over which π is strictly increasing, condition $c_m(\theta_1, \theta_2, \pi, \lambda = 1)$ holds, and the associated wedge is non-negative.

The Lagrange multiplier must be constant above θ_n because the rule burns money, and it must be equal to 1 because $\Lambda(\bar{\theta}) = 1$. If the monotonicity condition (11) in condition c_m does not hold over a subset of $[\theta_n, \bar{\theta}]$, then the rule may feature bunching. Proposition 2 implies that the following condition must hold over any interval of shocks over which there is bunching and money burning,

(c2) let $(\theta_1, \theta_2) \subset (\theta_n, \bar{\theta}]$ denote an interval over which π is constant and $\pi(\theta) \neq \pi(\theta_1)$ for $\theta \notin [\theta_1, \theta_2]$, condition $c_b(\theta_1, \theta_2, \pi(\theta_1), \underline{\lambda} = 1, \bar{\lambda} = 1)$ holds.

Again, the Lagrange multiplier must be constant above θ_n because the rule burns money, and it must be equal to 1 because $\Lambda(\bar{\theta}) = 1$.

For shock realizations below θ_n, Proposition 2 implies that the following condition must hold for the bunching of the policy schedule over $[\theta_x, \theta_n]$ in which case :

(c3) if $\theta_x < \theta_n$, then condition $c_b(\theta_x, \theta_n, \pi(\theta_x), \underline{\lambda} = 0, \bar{\lambda} = 1)$ holds.

The lower value of the Lagrange multiplier step function is necessarily 0 if $\theta_x = \underline{\theta}$ (by Proposition 2). The value of $\underline{\lambda}$ does not matter if $\theta_x > \underline{\theta}$ because in that case, by definition of a graduated schedule of penalties, $\pi^* = \pi_f(\theta_x)$ in inequality (12), and the step function jumps at θ_x where $\Delta(\theta_x, \pi^*) = 0$. The higher value of the Lagrange multiplier step function is necessarily 1 because the Lagrange multiplier, which is right-continuous, must be constant above θ_n because the rule burns money. It must be equal to 1 because $\Lambda(\bar{\theta}) = 1$.

If the policy can respond flexibly to shocks between $[\underline{\theta}, \theta_x]$, then Proposition 3 indicates that the following condition must hold:

(c4) if $\underline{\theta} < \theta_x$, then $\nu_\pi(\underline{\theta}, \pi_f(\underline{\theta})) \geq 0$ and condition $c_d(\underline{\theta}, \theta_x)$ holds.

The next proposition shows that the necessary optimality conditions (c1)-(c4) are jointly sufficient for the optimality of a rule that imposes a graduated schedule of penalties.

Proposition 4 (Optimality of rules imposing a graduated schedule of penalties)**.** *Suppose $\rho < \kappa$ and let (π, τ) denote a graduated schedule of penalties parametrized by $\theta_n \in (\underline{\theta}, \bar{\theta})$, and $\theta_x \in [\underline{\theta}, \theta_n]$. The rule (π, τ) is optimal if and only if (c1), (c2), (c3), and (c4) hold, and τ satisfies the envelope condition (4).*

The proof is in Appendix A.6.

Application to the design of fiscal rules. Continuing with the previous application, suppose that $\beta = \rho$ so that the monotonicity condition for marginal deterrence is equivalent to log-convexity of the tail $1 - F$. In contrast to the previous application, suppose that the distribution of shocks has a log-convex tail $1 - F$ up to a threshold above which the tail is log-concave (e.g., the truncated Pareto distribution).[13] Unlike the log-concave case of the previous application—for which marginal deterrence is never optimal and the fiscal rule is simply a cap on spending—a log-convex tail makes marginal deterrence a valid building block of an optimal fiscal rule.

A main takeaway from this application is that the policy schedule that satisfies the wedge with $\lambda = 1$,

$$\Delta(\theta, \pi(\theta)) = \frac{1}{1-\rho}\left((1-\beta)\theta - \rho\frac{1-F(\theta)}{f(\theta)}\right), \tag{14}$$

is useful in designing an optimal fiscal rule. Proposition 1 implies that the optimality of marginal deterrence necessarily implements policies that satisfy this wedge. In addition, marginal deterrence can only be part of an optimal rule over the range of shocks for which the implemented public spending schedule is non-decreasing.

An optimal fiscal rule, however, cannot feature marginal deterrence at the top. If a graduated schedule of penalties on public spending is an optimal fiscal rule, because the tail is log-concave above a threshold, the optimal fiscal rule also imposes a cap on public spending. To see this, suppose by way of contradiction that the optimal fiscal rule is a graduated schedule of penalties on public spending that does not impose a cap. Because the rule burns money above a threshold

[13] Truncating the tail to a compact support implies a log-convex tail up to a threshold above which the tail is log-concave.

and the implemented public spending schedule is monotonic, if there is no cap, it must be that the public spending schedule is strictly increasing and satisfies the wedge with $\lambda = 1$. The monotonicity condition, however, implies that the public spending schedule associated with the wedge with $\lambda = 1$ is decreasing at the top (since the tail $1 - F$ is log-concave tail near the top of the support), which contradicts incentive compatibility of the rule. The log-concavity of the tail of the distribution of shocks above a threshold implies that the necessary optimality condition (c2) is satisfied for $\theta_1 = \theta_p$, and $\theta_2 = \bar{\theta}$ because the mean residual life of the distribution of shocks is decreasing above the threshold θ_p, which solves

$$(1 - \rho)\Delta(\theta_p, \pi(\theta_p)) = \theta_p - \beta E[\theta | \theta \geq \theta_p].$$

If a graduated schedule of penalties on public spending is an optimal fiscal rule, then it either grants (bounded) discretion to respond to shocks below a threshold, or it grants an exemption from marginal deterrence below a threshold. The sign of the wedge (14) evaluated at $\underline{\theta}$ serves as a guide to designing the rule at the bottom. If the wedge is negative at $\underline{\theta}$—as in the bottom left panel in Figure 3—then, because the tail is log-concave above the threshold, the wedge is strictly increasing there and hence positive above some $\theta_n \in [\underline{\theta}, \bar{\theta})$. For the distributions in Figure 3 the wedge single-crosses zero, so it is negative for $\theta \leq \theta_n$ and positive for $\theta > \theta_n$. By Proposition 3, granting discretion below θ_n can only be optimal if condition (c4) holds for $\theta_x = \theta_n$. If, instead, the wedge is strictly positive at the lowest possible shock realization—as in the bottom right panel in Figure 3—then the optimal rule cannot burn money in the neighborhood of $\underline{\theta}$. To see this formally, suppose by way of contradiction that the optimal rule burns money for $\theta > \theta_n = \underline{\theta}$. Complementary slackness then makes the Lagrange multiplier constant on $(\underline{\theta}, \bar{\theta}]$, so it jumps from $\Lambda(\underline{\theta}) = 0$ to $\Lambda(\underline{\theta}^+) = 1$. This jump contributes $\rho\Delta(\underline{\theta}, \pi(\underline{\theta}))$ to $\delta(\underline{\theta})$. This violates the boundary optimality condition $\delta(\underline{\theta}) = 0$ since the wedge is strictly positive at $\underline{\theta}$. If the optimal rule grants an exemption from marginal deterrence below a threshold, the resulting kink in the penalty schedule implements bunching of the policy schedule over an interval denoted $[\theta_x, \theta_n]$, which can only be optimal if condition (c3) holds.

Whether the rule grants discretion or an exemption depends on the severity of the bias relative to the incentive cost of marginal deterrence. The numerical examples depicted in the bottom row of Figure 3 illustrate the two cases. In the bottom right panel, the distribution of shocks follows a truncated Pareto distribution with tail parameter $\gamma = 4$ (the tail parameter is such that the expected fiscal needs are the same as that under the exponential distribution). The wedge is strictly positive throughout the range of possible shocks and the conditions for bunching at the bottom and at the top of the range of shocks are satisfied. In the bottom left panel, the distribution of shocks, denoted F_a, is defined by its hazard rate, which is a convex combination of

the hazard rates of the exponential distribution and the Pareto distribution, $h_a(\theta) = a\lambda + (1-a)\frac{\gamma}{\theta}$, with $a = 0.375$. The choice of a is such that the expected fiscal need is the same across all four panels. [14] The inverse hazard rate of the distribution F_a implies a marginal penalty that is increasing in severity, as seen by the widening gap between the discretionary policy schedule and the policy schedule associated with the wedge with $\lambda = 1$ in the bottom-left panel in Figure 3. Truncating the distribution F_a implies that there is a threshold below which the distribution is log-convex, and above which the distribution is log-concave, which implies a non-monotonic policy schedule associated with the wedge with $\lambda = 1$ in the bottom-left panel in Figure 3.

This application is useful to evaluate the Excessive Deficit Procedure of the Stability and Growth Pact. In theory, the fiscal rule imposes a flat fine on European nations whose deficits exceed 3% of GDP (in practice the rule has not been enforced).[15] The only rationale for a notch in the penalty schedule—here, a jump from zero penalty to 0.25 percent of GDP for deficits above 3% of GDP—is to implement a cap on deficits. A cap without marginal deterrence can be an optimal fiscal rule for a relatively thin tail of the distribution of fiscal needs (top row in Figure 3). In contrast, for a thicker tail of the distribution of fiscal needs, the optimal fiscal rule favors marginal deterrence (bottom row in Figure 3); that is a penalty that scales continuously with the excess deficit above a threshold.

5 Applications

The tools to solve a delegation problem with the possibility of burning money are useful beyond environments with a constant bias, as in the design of a fiscal rule. In this section, I study the design of a trade agreement in a model of international trade featuring privately observed shocks to the political pressure for protectionism and the regulation of a monopolist with unknown marginal costs. In each application, the cost of money burning reflects a distinct economic primitive—the inefficiency with which an importing country compensates its trading partners for protection, and the degree of regulatory capture. Varying the cost of money burning bridges

[14] The hazard rate h_a defines the untruncated distribution $F_a(\theta) = 1 - \exp\left(-\int_{\underline{\theta}}^{\theta} h_a(x)\,dx\right)$ on $[\underline{\theta}, \infty)$; the analysis uses its truncation to the compact support $[\underline{\theta}, \bar{\theta}]$, with density $f_a(\theta)/F_a(\bar{\theta})$ for $\theta \in [\underline{\theta}, \bar{\theta}]$.

[15] With limited enforcement, it is optimal to enforce the rule to the extent possible (Halac and Yared (2022a)), which was not the case in Europe. Even an unenforced rule can still discipline policy through ex post renegotiation (Piguillem and Riboni (2021)), and the scope for renegotiation substantially changes the optimal balance between discipline and discretion in the presence of political risk (Dovis, Kirpalani, and Sublet (2026)). Non-enforcement, however, also damages the reputation of the central authority, which introduces perverse incentives for local governments (Dovis and Kirpalani (2020)).

two polar cases in the literature: trade agreements with and without compensation (Bagwell and Staiger (2005) and Amador and Bagwell (2013)), and the regulation of a monopolist with and without transfers (Baron and Myerson (1982) and Alonso and Matouschek (2008)).

In each application the necessary and sufficient optimality conditions serve complementary purposes. The necessary conditions evaluate whether an existing institution—the tariff binding, tariff overhang, and safeguards of the WTO rule, or a price cap and graduated fines in the regulation of a monopolist—is consistent with an optimal rule. The sufficient conditions then construct an optimal rule and identify avenues for reform.

5.1 Designing a trade agreement

Following Bagwell and Staiger (2005), consider a two-country model with perfect competition and terms-of-trade externalities. Home and Foreign both produce the goods x and n where n denotes the numéraire. Consumers have quasi-linear utility $u(c_x) + c_n$, and $u(c) = c - c^2/2$. The numéraire is freely traded and produced, using labor, with constant returns to scale in both countries. The wage and the price of the numéraire are normalized to one in both countries. The supply of good x is $Q(p) = p/2$ in the Home country, and $Q_*(p_*) = p_*$ in Foreign, where p and p_* denote the price of good x in the respective countries. As a result, Home's demand for imports is $p(z) = \frac{2}{3}(1-z)$, and Foreign's supply of exports is $p_*(z) = \frac{1}{2}(1+z)$, where z denotes the volume of trade of good x. Home's exports of the numéraire ensure that trade is balanced. The Home country's trade policy is an import tariff on good x denoted t.

Import tariffs affect the terms of trade, the volume of trade, and profits. The equilibrium condition for the terms of trade, $p(z) = p_*(z) + t$ can be rewritten to show that import tariffs reduce the volume of trade:

$$z(t) = \frac{1}{7} - \frac{6}{7}t.$$

In turn, profits at Home depend negatively on the volume of trade,

$$\pi(z) \equiv \int_0^{p(z)} Q(\tilde{p})d\tilde{p} = \frac{(1-z)^2}{9}.$$

As expected, import tariffs increase profits at Home.

The Home government—the agent—maximizes a weighted sum of the producer and consumer surpluses in the Home country, $U^A(\theta, \pi) = \theta\pi + b(\pi)$, where the weight $\theta \in [\underline{\theta}, \bar{\theta}]$ denotes a shock to the political pressure for trade protection, and $b(\pi) = (9\sqrt{\pi} - 17\pi - 1)/2$ denotes the consumer surplus inclusive of tariff revenues.

The objective of the designer of the trade agreement—the principal—differs from that of the

Home government due to the terms-of-trade externality that Home's import tariffs impose on Foreign. The objective of the designer of the trade agreement is the sum of the welfare in the Home and the Foreign countries, $U^P(\theta,\pi) = U^A(\theta,\pi) + v(\pi)$ where v denotes the welfare in the Foreign country. The welfare of Foreign, which is the sum of Foreign consumer and producer surpluses, is decreasing in the import tariffs imposed by Home. Expressed as a function of profits in the Home country, $v(\pi) = (2 - 6\sqrt{\pi} + 9\pi)/4$.

Due to the terms-of-trade externality, the Home government's preferred tariff is inefficiently high. The bias, however, is decreasing in the political pressure for trade protection (i.e., formally, the objective of the principal is less concave than that of the agent $\kappa = 2/3$). Let $\bar{\theta} < 7/4$ so that the bias of the Home government is strictly positive (for $\theta \geq 7/4$, the political pressure on the Home government would be so large that the designer of the trade agreement and the Home government would agree that autarky is the preferred outcome). In addition, suppose that the range of shocks is sufficiently large in the following sense: $\bar{\theta} - \underline{\theta} > 2\left(\frac{7}{4} - \bar{\theta}\right)$.[16] A trade agreement is a schedule of domestic profits and money burning $(\pi(\theta), \tau(\theta))$. Money burning refers to the penalty that the trade agreement imposes on the Home country for tariffs that implement profits $\pi(\theta)$. An optimal trade agreement maximizes $\int_\Theta \left[U^P(\theta, \pi(\theta)) - \rho\tau(\theta)\right] dF(\theta)$ subject to the incentive compatibility constraints (2) and the money-burning constraint (3).

The cost of enforcement parameter ρ captures the inefficiency with which a penalty on the importing country compensates the exporting country for the tariff. The penalty can be interpreted literally as a transfer payment, in which case $1-\rho$ denotes the fraction of the transfer from Home that reaches the Foreign country (e.g., a fraction ρ of the transfer is lost to bureaucratic inefficiencies). Because direct transfers have never been part of the GATT/WTO rule, Bagwell and Staiger (2005) discuss alternative compensation mechanisms that have been used in practice. The GATT rule took a permissive stance on Voluntary Export Restraints, which amount to a transfer of the tariff revenue to the exporting country. Under the WTO rule, instead, the penalty for activating a safeguard measure is the lack of flexibility to reactivate the safeguards for a period of time equal to the duration of its prior use. In a dynamic setting, the Foreign country benefits from Home's lack of flexibility to reactivate the safeguards. As a result, the money burnt by a dynamic use constraint on safeguards asymmetrically affects the objective of the designer of the trade agreement and that of Home.[17]

[16] If the range of shocks does not satisfy the inequality, the best cap leaves no discretion.

[17] In Online Appendix OA3, I use the repeated game in Bagwell and Staiger (2005) to show that the cost of enforcement implied by a dynamic use constraint on safeguards is $\rho(\tau) \leq 2/3$. In the repeated game, the cost of enforcement is a function of money burning because the benefit to Foreign is a non-linear function of the penalty on Home.

First, I use the necessary optimality conditions to infer features of the environment that are consistent with the local features of an optimal rule. In particular, I infer features of the environment under which the current WTO rule is qualitatively consistent with an optimal rule in the model. The current design of the WTO rule specifies thresholds known as tariff bindings, below which members have discretion regarding their applied tariffs, and above which flexibility is conditional on the activation of escape clauses known as safeguards. Under the WTO rule, there is evidence of bunching at the tariff bindings, applied tariffs below the tariff bindings (which is known as tariff overhang), and safeguards are occasionally activated (Table 3 in Bagwell, Bown, and Staiger (2016), and Beshkar, Bond, and Rho (2015)).

If the optimal rule implements the three qualitative features of the WTO rule outlined above—bunching at a threshold, discretion below the threshold, and flexibility with money burning above the threshold, then the model must satisfy the necessary optimality conditions in Propositions 2, 3, and 1. Proposition 2 implies that an optimal trade agreement may induce the bunching of applied tariffs at a tariff binding only under a joint condition on inefficiency of enforcement ρ and the distribution of shocks.

Corollary 5 (Bunching at a tariff binding)**.** *Let (π, τ) be an optimal trade agreement. If $\pi(\theta) = \pi_f(\theta_1)$ for $\theta \in (\theta_1, \theta_2)$ and $\pi(\theta) \neq \pi_f(\theta_1)$ for $\theta \notin [\theta_1, \theta_2]$, then there exists $\lambda \in [0, 1]$ such that the tail of the distribution of shocks satisfies*

$$\theta_1 - E[\hat{\theta} | \theta_1 \leq \hat{\theta} \leq \theta] \leq v'(\pi_f(\theta_1)) + \rho(\theta - \theta_1) \frac{\lambda - F(\theta)}{F(\theta) - F(\theta_1)}$$

for $\theta \in (\theta_1, \theta_2)$, and as an equality for $\theta = \theta_2$. If $\theta_2 = \bar{\theta}$ (i.e., a tariff cap), or $\tau(\theta) > 0$ for $\theta > \theta_2$ (i.e., safeguards), then $\lambda = 1$.

For instance, for uniformly distributed shocks, the inequality condition in Corollary 5 implies that an optimal trade agreement may induce bunching at the tariff binding only if more than half of the compensation is wasted, i.e., $\rho \geq 1/2$. Similarly, Proposition 3 implies that an optimal trade agreement may grant discretion under a tariff binding only under a joint condition on the distribution of shocks and the cost of enforcement ρ.

Corollary 6 (Tariff overhang)**.** *Let (π, τ) be an optimal trade agreement that grants discretion below $\pi_f(\theta_1)$, then for $\theta \leq \theta_1$,*

$$\left(\frac{7}{4} - \theta \right) \frac{f'(\theta)}{f(\theta)} \geq 1 - 3\rho.$$

For instance, granting discretion to respond to uniformly distributed shocks can be part of an optimal trade agreement only if $\rho \geq 1/3$. The alternative to granting flexibility and bunching is

marginal deterrence. The next corollary to Proposition 1 shows that safeguards can be part of an optimal trade agreement only under a joint condition on the cost of enforcement ρ and a measure of the thickness of the right-tail of the distribution of shocks to the benefit of protectionism.

Corollary 7 (Safeguards)**.** *Let (π, τ) be an optimal trade agreement. If the trade agreement burns money on profits in $[\pi(\theta_1), \pi(\theta_2)]$ and π is strictly increasing over $[\theta_1, \theta_2]$, then there exists $\lambda \in [0, 1]$ such that*

$$\rho \frac{d}{d\theta} \frac{\lambda - F(\theta)}{f(\theta)} \geq \rho - 1,$$

and the marginal penalty is (10) with $\nu_\pi(\theta, \pi(\theta)) = -v'(\pi(\theta))$ for $\theta \in [\theta_1, \theta_2]$. If $\tau(\theta) > 0$ for $\theta > \theta_1$, then $\lambda = 1$.

For instance, for uniformly distributed shocks, an optimal trade agreement may feature safeguards only if $\rho \leq 1/2$, which, in turn, implies that $\lambda = 1$. To see that it implies that $\lambda = 1$, note that a sufficient condition for money burning to be strictly positive for $\theta > \theta_1$ is that money burning be non-decreasing for $\theta \geq \theta_2$, which is the case $\pi(\theta) \leq \pi_f(\theta)$ (that is, $\frac{7}{4} - \theta \geq 3\rho \frac{1-F(\theta)}{f(\theta)}$) for $\theta \in [\theta_2, \bar{\theta}]$.

Second, I construct an optimal rule based on the necessary optimality conditions, and use the sufficient conditions to verify that the guess is globally optimal. Equipped with a guess of the optimal rule based on the necessary optimality conditions, the sufficient optimality conditions allow one to verify that a rule that satisfies the necessary optimality conditions jointly is an optimal trade agreement. The necessary optimality conditions indicate that the optimal trade agreement depends on how efficiently trading partners can be compensated for protectionist tariff policies. For instance, under a uniform distribution, an optimal trade agreement cannot simultaneously feature tariff overhang, bunching at a tariff binding, and marginal deterrence (a combination that does arise in the data and, as panel d) shows, under a non-uniform distribution). The next corollary shows that if the enforcement mechanism is relatively inefficient, the optimal trade agreement is a tariff cap (as in Amador and Bagwell (2013)). Otherwise, the optimal trade agreement resembles a threshold with escape clauses enforced by a graduated schedule of penalties as a function of the surge in imports. Panels a), b), and c) in Figure 4 illustrate the next corollary to Corollary 4 on interval delegation and Proposition 4 on the optimality of a rule enforced by a graduated schedule of penalties.

Corollary 8 (Optimal trade agreement)**.** *For uniformly distributed shocks θ, there exists $\rho_1, \rho_2 \in (0, 1)$ such that the following holds.*

a) For $\rho_1 \leq \rho \leq \kappa$, the optimal trade agreement is a cap on tariffs.

b) For $\rho_2 \leq \rho < \rho_1$, the optimal trade agreement is a threshold with safeguards enforced by a differentiable penalty schedule so there is no bunching at the threshold.

c) For $0 < \rho < \rho_2$, the optimal trade agreement is a threshold with safeguards implemented by a penalty schedule that has a kink at the threshold and is differentiable above the threshold. Although there is bunching at the threshold, there is no tariff overhang.

Part a) of Corollary 8 corresponds to a result in Amador and Bagwell (2013). In contrast, if compensations are not too inefficient, the threshold with escape clauses as a function of observables—such as tariffs—is an optimal trade agreement. As the panels a) and c) in Figure 4 illustrate, if the optimal trade agreement features bunching at the tariff binding and shocks are uniformly distributed and privately observed, then tariff overhang and safeguards are mutually exclusive features of an optimal trade agreement.[18] This is, however, specific to economies with uniformly distributed shocks because optimal incentives are either bang-bang (for $\rho \geq 1/2$) or compatible with safeguards (for $\rho < 1/2$).

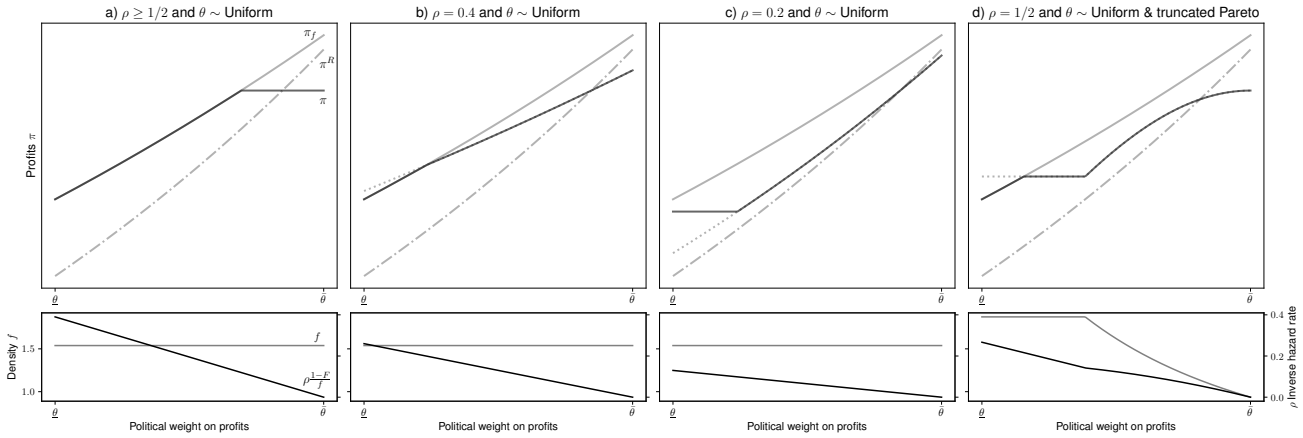

Figure 4: Optimal trade agreements.

Notes: Top panels: The solid-grey line depicts the Home government's preferred schedule of profits in response to political pressure ranging from $\underline{\theta} = 1$, which corresponds to equal weights on domestic consumer and producer surplus, to $\bar{\theta} = 1.65$. The dashed-grey line depicts the schedule of profits implemented by the Ramsey rule (i.e., the optimal rule if realized political pressure is public information). The solid-black line depicts the optimal rule with privately observed political pressure. The dotted line depicts the schedule of profits implemented by the marginal penalty (10) with $\lambda = 1$. Bottom panels: the grey line depicts the density of the distribution and the black line depicts the associated inverse hazard rate scaled by the cost of enforcement ρ.

The rich pattern of tariff policies observed in the data (i.e., tariff overhang, bunching at the tariff binding, and use of safeguards) obtains as an outcome of an optimal trade agreement in panel d) of Figure 4). With privately observed shocks, the inverse hazard rate of the distribution

[18] Interestingly, a similar result holds in a model with costly state verification, as shown in Proposition 4 (ii) in Beshkar and Bond (2017).

of shocks governs the incentive cost of flexibility. The distribution of shocks in panel d) of Figure 4) is Uniform up to a threshold above which it is truncated Pareto (i.e., spliced Uniform-truncated Pareto). The incentive cost of flexibility is high and steeply decreasing over the support of the distribution where shocks are Uniformly distributed, which leads to tariff overhang and bunching at the tariff binding. The incentive cost of flexibility decreases at a lower rate over the support of the truncated Pareto distribution, which implies a higher need for flexibility. As a result, the optimal trade agreement implements an allocation that features tariff overhang, bunching at the tariff binding, and upward flexibility enforced by safeguards.

The findings complement previous findings that lend support to the design of the WTO. Bagwell and Staiger (2005) show that tariff bindings and safeguards, enforced by a dynamic use constraint, can be desirable features of a trade agreement in an environment with privately observed shocks to the benefit of protectionism. Amador and Bagwell (2013), however, provide conditions under which a tariff cap—without safeguards—is optimal among all mechanisms. Beshkar and Bond (2017) show that a combination of tariff bindings and safeguards conditional on the costly verification of the state are desirable. Halac and Yared (2020a) go further by providing conditions under which a threshold and escape clause is optimal among all mechanisms in a general environment with costly state verification.

Although the predictions of the model are qualitatively consistent with the data on applied tariffs under WTO rule, the analysis suggests avenues for reform. Instead of a dynamic use constraint that is only a function of the duration of the activation of the escape clause, the model suggests that making the graduated penalty schedule depend on the applied tariff in excess of the tariff binding would improve upon the design of the trade agreement.

5.2 Regulating a monopolist with unknown costs

Following Baron and Myerson (1982), consider the regulation of a monopolist with unknown marginal costs. Let p denote the inverse demand function with constant elasticity $\epsilon > 0$.[19]The monopolist—i.e., the agent—produces at a constant marginal cost $\theta \in [\underline{\theta}, \bar{\theta}]$, which is private information to the monopolist. The revenue from selling π units is $b(\pi) = p(\pi)\pi$. The profits are $U^A(\theta, \pi) = -\theta\pi + b(\pi)$. The regulator—i.e., the principal—is uncertain about the marginal cost of the monopolist, with beliefs given by the distribution F. The regulator can tax the

[19]Although the analysis applies to a broader class of demand functions, the assumption of a constant elasticity is useful because it implies that the bias between the preferred regulated price of the principal and that of the agent is an increasing function of the privately observed shock (unlike the applications to the design of a fiscal rule and to the design of a trade agreement where the bias is constant and decreasing respectively).

monopolist, and tax revenues are rebated to the consumers. By the revelation principle, the tax policy $\tau(\theta)$ must be incentive compatible.

There are two political economy frictions. First, there is regulatory capture in the sense that the regulator places more weight on the surplus of the monopolist than that of the consumer, that is $\alpha \geq 1$ in[20]

$$U^P(\theta, \pi) = U^A(\theta, \pi) + \frac{1}{\alpha} v(\pi),$$

where $v(\pi)$ denotes the consumer surplus[21]

$$v(\pi) = \int_0^\pi p(\tilde{\pi}) d\tilde{\pi} - p(\pi)\pi.$$

Second, the money-burning constraints $\tau(\theta) \geq 0$ limit transfers from consumers to the monopolist. The money-burning constraint can be interpreted as society's effort to limit regulatory capture. In the absence of the money-burning constraint, a regulator subject to regulatory capture would maximize the transfer from consumers to the monopolist. In the Baron and Myerson (1982) model, instead, the regulator places less weight on the surplus of the monopolist than that of the consumer, i.e., $\alpha \leq 1$, and an individual rationality constraint on the part of the monopolist limits the transfer from the monopolist to the consumers.[22] For the limit case $\alpha = 1$, the regulator maximizes the total surplus and, as I show below, the optimal regulated price coincides with that in Baron and Myerson (1982).

The design of an optimal regulation maps to the mechanism design problem in Section 1 with $\rho = 1 - 1/\alpha$. [23] The regulator maximizes the weighted average of producer and consumer surpluses U^P net of the tax on the monopolist and including the lump-sum rebate of tax revenues

[20]The focus here is on the design of regulation by an agency that is subject to an exogenous form of regulatory capture. Chapter 11 in Laffont and Tirole (1993) offers a model of endogenous regulatory capture in which a third party—the legislature—provides incentives to the regulatory authority. Dal Bó (2006) offers a review of the literature.

[21]The objective of the regulator is more concave than that of the firm (i.e., $v'' < 0$) for a demand with constant elasticity.

[22]In this application, although the monopolist is assumed to operate to abstract from participation constraints, note that the optimal regulation implement positive ex post profits because fixed costs are assumed to be zero and the regulated price is above marginal costs (see Figure 5). Amador and Bagwell (2022) extend the applicability of Lagrangian methods to account for an ex post participation constraint in the design of an optimal regulation.

[23]Because $U^A_{\theta\pi} = -1 < 0$, output is non-increasing in the marginal cost, so the map to Section 1 reverses the order of types (equivalently, relabel $\tilde{\theta} = \underline{\theta} + \bar{\theta} - \theta$, which recovers the baseline form $\tilde{\theta}\pi + \tilde{b}(\pi)$ with $\tilde{b}$ strictly concave and leaves ρ and κ unchanged). The figures report the regulated price $p(\pi(\theta))$, which increases in the marginal cost θ.

to the consumers with welfare weight $1/\alpha$,

$$\int_{\Theta} \left[U^P(\theta, \pi(\theta)) - \tau(\theta) + \frac{1}{\alpha}\tau(\theta) \right] dF(\theta),$$

subject to the incentive compatibility constraint (2) and the money-burning constraint (3). For $\alpha = 1$, the regulator maximizes the gross surplus and, as a result, there is no cost to burning money because taxes on the monopolist, rebated to consumers, are neutral on welfare (i.e., $\rho = 0$). As $\alpha \to \infty$, the regulator places increasing weight on producer surplus and the cost of burning money $\rho \to 1$.

The necessary and sufficient optimality conditions make it possible to approach the problem from two complementary angles. First, I use the necessary optimality conditions to infer local optimality conditions. Specifically, I infer the subset of environments—i.e., restrictions on the degree of regulatory capture and the distribution of marginal costs—that are compatible with the optimality of local features of a regulation.

Corollary 9 (Price cap)**.** *Let (π, τ) be an optimal regulation. If there is a price cap at $p(\pi(\theta_p))$ for $\theta_p \in (\underline{\theta}, \bar{\theta})$, then $\alpha\big(E\big[\hat{\theta}\big|\hat{\theta} \geq \theta\big] - \theta\big) + \theta \leq \alpha\big(E\big[\hat{\theta}\big|\hat{\theta} \geq \theta_p\big] - \theta_p\big) + \theta_p$ for $\theta \in [\theta_p, \bar{\theta}]$. The threshold θ_p satisfies $p(\pi(\theta_p)) = \alpha\left(E[\theta|\theta \geq \theta_p] - \theta_p\right) + \theta_p$.*

For instance, for uniformly distributed marginal costs, a price cap is consistent with the objective of a regulator who places at least twice as much weight on producer surplus relative to consumer surplus. To see this, recall that the mean residual life of a uniform distribution has slope $-1/2$, hence the inequality in Corollary 9 can only be satisfied for $\alpha \geq 2$. In addition, data on the stringency of the cap and the threshold marginal cost at which the cap binds imply a joint condition on the degree of regulatory capture α and the conditional tail expectation of the distribution of shocks above the threshold marginal cost. If the regulation does not feature marginal deterrence, the threshold marginal cost can be inferred from the data on the stringency of the cap and the elasticity of demand.

If the regulation deters monopoly pricing at the margin instead, then the necessary optimality conditions determine the path of the optimal regulated price as a function of the Lagrange multiplier on the money-burning constraint and of a measure of the thickness of the distribution of shocks.

Corollary 10 (Marginal deterrence)**.** *Let (π, τ) be an optimal regulation. If the regulation levies taxes on output in $[\pi(\theta_2), \pi(\theta_1)]$, for $\theta_1 < \theta_2$, then there exists $\lambda \in [0, 1]$ such that, for $\theta \in [\theta_1, \theta_2]$, the regulated price is*

$$p(\theta) = \theta + (\alpha - 1)\frac{\lambda - F(\theta)}{f(\theta)},$$

and

$$(\alpha - 1)\frac{d}{d\theta}\frac{\lambda - F(\theta)}{f(\theta)} \geq -1.$$

If $\tau(\theta) > 0$ *for* $\theta \geq \theta_1$*, then* $\lambda = 1$*.*

For $\alpha = 1$, the optimal regulation implements the same price schedule as a regulator maximizing the gross surplus in the Baron and Myerson (1982) environment, and the distribution of the surplus between the monopolist and consumers differ only due to the different constraints on transfers. For $\alpha > 1$, however, the formula for the optimal regulation differs from that in Baron and Myerson (1982) because the incentive cost of marginal deterrence differs across the two environments. In the Baron and Myerson (1982) model, the individual rationality constraint binds for the transfer from the monopolist with the highest cost realization $\bar{\theta}$. As a result, the incentive cost of marginal deterrence is an increasing function of the left tail of the distribution of marginal costs (i.e., the regulated price in Baron and Myerson (1982) corresponds to $\lambda = 0$). In contrast, for $\alpha > 1$, the money-burning constraint binds for the transfer to the monopolist with low cost realizations and the incentive cost of marginal deterrence is an increasing function of the right tail of the distribution of marginal costs if the money-burning constraint does not bind for $\theta \geq \theta_1$, in which case $\lambda = 1$. A sufficient (but not necessary) condition for $\lambda = 1$ is that the money burning schedule is non-decreasing, which is the case if the wedge is non-negative; formally, if $\frac{1}{(\alpha-1)(\epsilon-1)} > \frac{1-F(\theta)}{\theta f(\theta)}$ for $\theta \in [\theta_1, \bar{\theta}]$, then $\lambda = 1$.

For uniformly distributed shocks, the monotonicity condition (11) indicates that levying a graduated tax schedule on the monopolist can only be part of an optimal regulation if $\alpha < 2$. Hence, for a uniform distribution, an optimal regulation can either feature marginal deterrence (for $\alpha < 2$), or a price cap (for $\alpha \geq 2$), but not both.

Corollary 11 (Flexibility)**.** *Let* (π, τ) *be an optimal regulation that grants discretion to implement output in* $[\pi_f(\theta_2), \pi_f(\theta_1)]$ *for* $\theta_1 < \theta_2$*, then for* $\theta \in [\theta_1, \theta_2]$*,*

$$\theta\frac{f'(\theta)}{f(\theta)} \geq -(\alpha - 1)(\epsilon - 1) - 1.$$

This condition is satisfied for uniformly distributed marginal costs.

As the degree of regulatory capture increases, the optimal regulation tends towards "bang-bang" incentives. In the absence of regulatory capture, the regulated price is equal to the marginal cost. In contrast, for a sufficiently large degree of regulatory capture, the regulation does not impose any discipline (as α increases, the Ramsey regulated price converges pointwise to the flexible price schedule). In between the polar cases of first-best regulation with no regulatory capture and no regulation with extreme regulatory capture, the degree of progressivity of the

regulation depends on the thickness of the right-tail of the distribution of marginal costs (see Figure 5).

Intuitively, there are two ways to avoid burning money: prohibitively large fines or no fines.[24] Figure 5 shows that the best way to avoid burning money depends on the thickness of the tail of the distribution of shocks. If the right-tail is log-concave, then the schedule of regulated prices flattens as the degree of regulatory capture increases. In contrast, if the right-tail is log-convex, then the schedule of regulated prices weakly steepens as the degree of regulatory capture increases. Figure 5 depicts the comparative statics of the regulated price with respect to the degree of regulatory capture α for uniformly distributed marginal costs in panel a), and for Pareto distributed marginal costs in panel b).[25]

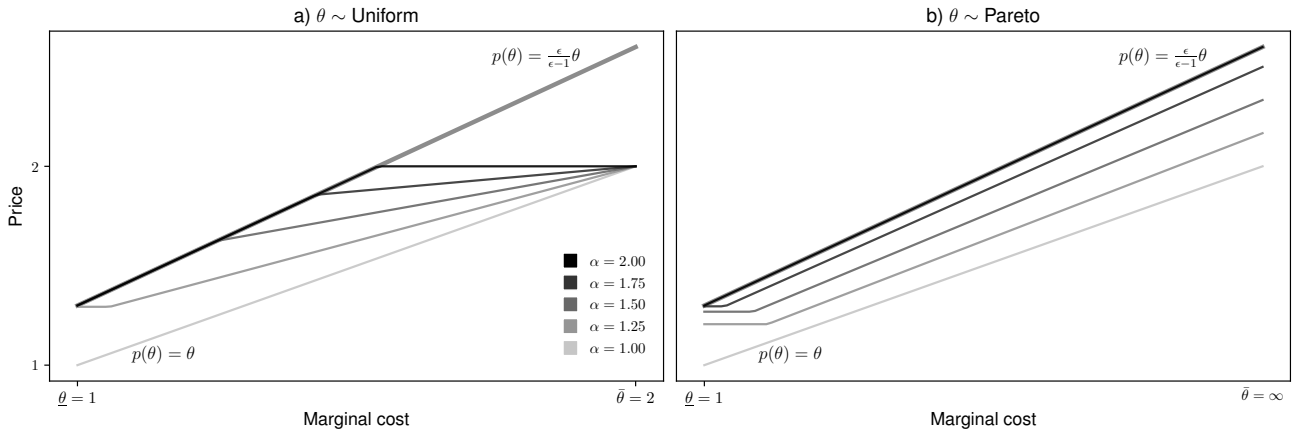

Figure 5: Regulated monopoly price as a function of regulatory capture α

Notes: For both panels, the elasticity of demand is constant at $\epsilon = 4.3\overline{3}$ to target a markup of 1.3. The highest line depicts the flexible price schedule, which is a constant markup over the marginal cost. The lowest line, depicting price equals marginal cost, corresponds to the regulated price in the absence of regulatory capture (i.e., $\alpha = 1$). In between the highest and lowest lines, darker lines depict the regulated price for higher degrees of regulatory capture α. Although the expected marginal cost is the same across both panels, the distribution of marginal costs is Uniform[1,2] for the left panel, and Pareto with tail parameter $\gamma = 3$ for the right panel.

The outcome of the theory is consistent with differences in monopoly regulation and anti-competitive practices in Europe and the United States. In the model, regulatory capture—measured by the excess weight that the regulator places on the producer surplus—induces leniency over fines. Since lobbying is more prominent in the U.S. than in Europe, it is plausible

[24]Formally, at any point of differentiability of the money burning schedule, $\tau'(\theta) = \Delta(\theta, \pi(\theta))\pi'(\theta)$. Prohibitive penalties set $\pi'(\theta) = 0$, whereas not imposing discipline sets $\Delta(\theta, \pi(\theta)) = 0$.

[25]The Pareto distribution has unbounded support, i.e., $\bar{\theta} = \infty$, whereas the model assumes $\bar{\theta} < \infty$. The sufficiency results illustrated in Figure 5 carry over to environments with $\bar{\theta} = \infty$, provided the first moment of the distribution is finite, as shown in Sublet (2026).

that U.S. regulators place a relatively larger weight on producer surplus than their European counterparts do.[26] Consistent with the prediction of the theory, U.S. antitrust doctrine takes a lenient stance on excessive pricing. For instance, the application of the Sherman Act paragraph 2 in Verizon v. Trinko (2004) does not penalize excessive pricing (instead, it penalizes exclusionary anti-competitive practices). In contrast, European competition law penalizes excessive pricing with fines (based on the interpretation of Article 102 TFEU in the rulings United Brands v. Commission (1978) and Aspen Pharma (2020)).

6 Conclusion

I studied the optimal design of a policy rule in a general principal-agent model of the trade-off between discipline and flexibility. By extending the applicability of global Lagrangian methods, I obtain necessary and sufficient optimality conditions that completely characterize optimal rules. The necessary optimality conditions hold for any optimal rule, including rules with discontinuities. Three features of the model determine the optimal compromise between discipline and flexibility: the bias, the thickness of the tail of the distribution of shocks, and the principal's aversion to burning money. The analysis bridges the gap between two polar cases for the principal's aversion to burning money. At one end of the spectrum, the principal is immune to burning money, in which case the optimal rule resembles a Pigouvian tax. At the opposite end of the spectrum, money burning is symmetric in that penalties are joint for the principal and the agent, and the optimal rule features bang-bang incentives. In between, for a thick tail, an optimal rule features marginal deterrence at the cost of burning money, whereas for a thin tail, an optimal rule limits the choice set.

The global Lagrangian method is widely applicable to solve mechanism design problems with limited transfers in other contexts. Werning (2007) studies Pareto efficient income taxation, where the requirement that a reform be Pareto improving limits transfers. The method also applies to designing the cost of verifying the state to determine whether escape clauses apply (see Halac and Yared (2020a) for the design of rules with costly state verification). Another promising application is the study of the optimal illiquidity of retirement savings accounts for households who undersave for their retirement (see Laibson et al. (1998) and Beshears et al. (2020)).

[26] Grossman and Helpman (1994) provides a theory mapping lobbying expenses to welfare weights in the regulator's objective function, and Gutiérrez and Philippon (2023) contains a comparison of lobbying expenses by firms in the U.S. and Europe.

An interesting direction for future work is to explore the optimality of dynamic mechanisms in repeated games in which penalties arise endogenously through strategic behavior. On-equilibrium penalties in such settings are analogous to burning money. Symmetric money burning arises naturally in several repeated-game environments (Athey, Bagwell, and Sanchirico (2004), Athey, Atkeson, and Kehoe (2005), and Halac and Yared (2025)), where the genericity of bang-bang incentives makes static mechanisms optimal. By contrast, asymmetric money burning arises naturally—as I show in Appendix OA3—in the international trade environment of Bagwell and Staiger (2005) and in the regulation of a monopolist (Section 5).

Another interesting avenue for future work is the quantification of the trade-off between discipline and discretion. Kang and Silveira (2021) quantify the benefit of discretion in the enforcement of environmental regulation in a fully specified structural model. Based on the theory in this paper, the thickness of the tail of the distribution of shocks to fiscal needs is a quantifiable measure of the need for flexible fiscal rules. In Sublet (2026), I infer past fiscal needs from government finance data through the lens of the model of fiscal policy.

References

Akbarpour, M., Dworczak, P., and Kominers, S. D. (2024). Redistributive Allocation Mechanisms. Journal of Political Economy, 132(6): 1831-1875

Alonso, R., and Matouschek, N. (2008). Optimal Delegation. Review of Economic Studies, 75: 259-293

Amador, M., and Bagwell, K. (2013). The Theory of Optimal Delegation With an Application to Tariff Caps. Econometrica, 81: 1541-1599

Amador, M., and Bagwell, K. (2020). Money Burning in the Theory of Delegation. Games and Economic Behavior, 112: 382-412

Amador, M., and Bagwell, K. (2022). Regulating a Monopolist With Uncertain Costs Without Transfers. Theoretical Economics, 17(4): 1719-1760

Amador, M., Bagwell, K., and Carpizo, V. (2026). Generalized Delegation. Working paper.

Amador, M., Werning, I., and Angeletos, G.-M. (2006). Commitment vs. Flexibility. Econometrica, 74: 365-396

Ambrus, A., and Egorov, G. (2017). Delegation and Nonmonetary Incentives. Journal of Economic Theory, 171: 101-135

Athey, S., Atkeson, A., and Kehoe, P. (2005). The Optimal Degree of Discretion in Monetary Policy. Econometrica, 73(5): 1431-1475

Athey, S., Bagwell, K., and Sanchirico, C. (2004). Collusion and Price Rigidity. Review of Economic Studies, 71(2): 317-349

Atkeson, A., and Lucas, R. E. (1992). On Efficient Distribution with Private Information. Review of Economic Studies, 59(3): 427-453

Bagwell, K., Bown, C. P., and Staiger, R. W. (2016). Is the WTO Passé? Journal of Economic Literature, 54(4): 1125-1231

Bagwell, K., and Staiger, R. W. (2005). Enforcement, Private Political Pressure, and the GATT/WTO Escape Clause. Journal of Legal Studies, 34(2): 471-513

Baron, D. P., and Myerson, R. B. (1982). Regulating a Monopolist with Unknown Costs. Econometrica, 50(4): 911-930

Beshears, J., Choi, J. J., Clayton, C., Harris, C., Laibson, D., and Madrian, B. C. (2020). Optimal Illiquidity. Working paper.

Beshkar, M., and Bond, E. W. (2017). Cap and Escape in Trade Agreements. American Economic Journal: Microeconomics, 9(4): 171-202

Beshkar, M., Bond, E. W., and Rho, Y. (2015). Tariff Binding and Overhang: Theory and Evidence. Journal of International Economics, 97(1): 1-13

Boerma, J., Tsyvinski, A., and Zimin, A. P. (2025). Bunching and Taxing Multidimensional Skills. Working paper.

Carter, M., and van Brunt, B. (2000). *The Lebesgue-Stieltjes Integral: A Practical Introduction.* New York: Springer-Verlag

Clayton, C., and Schaab, A. (2025). A Theory of Dynamic Inflation Targets. American Economic Review, 115(2): 448-490

Dal Bó, E. (2006). Regulatory Capture: A Review. Oxford Review of Economic Policy, 22(2): 203-225

Diamond, P. (1998). Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates. American Economic Review, 88: 83-95

Dovis, A., and Kirpalani, R. (2020). Fiscal Rules, Bailouts, and Reputation in Federal Governments. American Economic Review, 110(3): 860-888

Dovis, A., and Kirpalani, R. (2021). Rules without Commitment: Reputation and Incentives. Review of Economic Studies, 88(6): 2833-2856

Dovis, A., Kirpalani, R., and Sublet, G. (2026). The Optimal Allocation of Policy-Making Between Executive Agencies and the Legislature. Working paper.

Galperti, S. (2015). Commitment, Flexibility, and Optimal Screening of Time Inconsistency. Econometrica, 83: 1425-1465

Goltsman, M., Hörner, J., Pavlov, G., and Squintani, F. (2009). Mediation, Arbitration and Negotiation. Journal of Economic Theory, 144(4): 1397-1420

Grossman, G. M., and Helpman, E. (1994). Protection for Sale. American Economic Review, 84(4): 833-850

Gutiérrez, G., and Philippon, T. (2023). How European Markets Became Free: A Study of Institutional Drift. Journal of the European Economic Association, 21(1): 251-292

Halac, M., and Yared, P. (2014). Fiscal Rules and Discretion under Persistent Shocks. Econometrica, 82: 1557-1614

Halac, M., and Yared, P. (2018). Fiscal Rules and Discretion in a World Economy. American Economic Review, 108: 2305-2334

Halac, M., and Yared, P. (2020a). Commitment vs. Flexibility with Costly Verification. Journal of Political Economy, 128: 4523-4573

Halac, M., and Yared, P. (2020b). Inflation Targeting under Political Pressure. In Independence, Credibility, and Communication of Central Banking, edited by Ernesto Pastén and Ricardo Reis, Santiago, Chile: Central Bank of Chile, 7-27

Halac, M., and Yared, P. (2022a). Fiscal Rules and Discretion under Limited Enforcement. Econometrica, 90: 2093-2127

Halac, M., and Yared, P. (2022b). Instrument vs. Target-Based Rules. Review of Economic Studies, 89: 312-345

Halac, M., and Yared, P. (2025). A Theory of Fiscal Responsibility and Irresponsibility. Journal of Political Economy, 133(5): 1574-1620

Holmström, B. (1977). On Incentives and Control in Organizations. Ph.D. thesis, Stanford Graduate School of Business

Kang, K., and Silveira, B. (2021). Understanding Disparities in Punishment: Regulator Preferences and Expertise. Journal of Political Economy, 129(10): 2947-2992

Kartik, N., Kleiner, A., and Van Weelden, R. (2021). Delegation in Veto Bargaining. American Economic Review, 111(12): 4046-4087

Kleiner, A., Moldovanu, B., and Strack, P. (2021). Extreme Points and Majorization: Economic Applications. Econometrica, 89(4): 1557-1593

Kolotilin, A., and Zapechelnyuk, A. (2025). Persuasion Meets Delegation. Econometrica 93(1): 195-228

Kováč, E., and Mylovanov, T. (2009). Stochastic Mechanisms in Settings without Monetary Transfers: The Regular Case. Journal of Economic Theory, 144: 1373-1395

Laffont, J.-J., and Tirole, J. (1993). A Theory of Incentives in Procurement and Regulation. Cambridge, MA: MIT Press

Laibson, D., Repetto, A., Tobacman, J., Hall, R., Gale, W., and Akerlof, G. A. (1998). Self-Control and Saving for Retirement. Brookings Papers on Economic Activity, 91-196

Luenberger, D. G. (1969). *Optimization by Vector Space Methods.* Series in Decision and Control. New York: Wiley

McKibbin, W. J., and Wilcoxen, P. J. (2002). The Role of Economics in Climate Change Policy. Journal of Economic Perspectives, 16(2): 107-129

Melumad, N. D., and Shibano, T. (1991). Communication in Settings with No Transfers. The Rand Journal of Economics, 22: 173-198

Mookherjee, D., and Png, I. P. L. (1994). Marginal Deterrence in Enforcement of Law. Journal of Political Economy, 102(5): 1039-1066

Myerson, R. B. (1981). Optimal Auction Design. Mathematics of Operations Research, 6: 58-73

Nöldeke, G., and Samuelson, L. (2007). Optimal Bunching without Optimal Control. Journal of Economic Theory, 134(1): 405-420

Persson, T., and Tabellini, G. (1993). Designing Institutions for Monetary Stability. Carnegie-Rochester Conference Series on Public Policy, 39: 53-84

Piguillem, F., and Riboni, A. (2021). Fiscal Rules as Bargaining Chips. Review of Economic Studies, 88(5): 2439-2478

Piguillem, F., and Riboni, A. (2024). Sticky Spending, Sequestration, and Government Debt. American Economic Review, 114(11): 3513-3550

Piguillem, F., and Schneider, A. (2016). Coordination, Efficiency and Policy Discretion. Working paper.

Rochet, J.-C., and Choné, P. (1998). Ironing, Sweeping, and Multidimensional Screening. Econometrica, 66(4): 783-826

Sublet, G. (2026). The Optimal Degree of Flexibility in Fiscal Policy: A Quantitative Exploration. Working paper.

Toikka, J. (2011). Ironing without Control. Journal of Economic Theory, 146(6): 2510-2526

Walsh, C. E. (1995). Optimal Contracts for Central Bankers. American Economic Review, 85(1): 150-167

Weitzman, M. (1974). Prices vs. Quantities. Review of Economic Studies, 41(4): 477-491

Werning, I. (2007). Pareto Efficient Income Taxation. Working paper.

Supplemental Appendix

A Proofs

This appendix contains the proofs of the main results (Lemmas 1 and 2, and Propositions 1–4).

A.1 Proof of Lemma 1

A.1.1 Necessity

The proof that the conditions are necessary proceeds in two steps. A first step shows that there exists a Lagrange multiplier function such that the Lagrangian is maximized at the solution and the complementary slackness condition holds. A second step shows that the optimality conditions in terms of Gateaux derivatives are necessary given that the solution is a maximizer of the Lagrangian.

The Lagrange multiplier function obtains as a supporting hyperplane to the primal functional. The primal functional results from imbedding the program of interest in a family of programs parametrized by the constraint on money burning. Let $X = \{\pi, \underline{\tau} \mid \pi : \Theta \mapsto \mathbb{R},\ \underline{\tau} \in \mathbb{R}\}$ and $\Omega = \{(\pi, \underline{\tau}) \in X \mid \pi \text{ is non-decreasing}\}$. Note that Ω is a convex cone and a subset of the linear vector space X. Based on (5), define the objective to be maximized as the functional $\hat{f} : \Omega \mapsto \mathbb{R}$:[27]

$$\begin{aligned}\hat{f}(\pi, \underline{\tau}) = \int_{\Theta} \left[U^P(\theta, \pi(\theta)) - \rho U^A(\theta, \pi(\theta)) + \rho\pi(\theta)\frac{1 - F(\theta)}{f(\theta)}\right] f(\theta) d\theta \\ - \rho\big(\underline{\tau} - U^A(\underline{\theta}, \pi(\underline{\theta}))\big).\end{aligned}$$

Let $G : \Omega \mapsto Z$ denote the money burning schedule,

$$G(\pi, \underline{\tau})(\theta) = \underline{\tau} - U^A(\underline{\theta}, \pi(\underline{\theta})) + U^A(\theta, \pi(\theta)) - \int_{\underline{\theta}}^{\theta} \pi(\tilde{\theta})\ d\tilde{\theta}.$$

where $Z = \{z \mid z : \Theta \mapsto \mathbb{R} \text{ is of bounded variation}\}$ with norm $||z|| = \sup_{\theta \in \Theta} |z(\theta)|$. Note that $G(\pi, \underline{\tau})$ is of bounded variation because π is non-decreasing and bounded for $(\pi, \underline{\tau}) \in \Omega$, b is strictly concave, and hence $\theta \mapsto \theta\pi(\theta) + b(\pi(\theta))$ can be expressed as the difference of two

[27] Following Amador and Bagwell (2013, p. 1579), let U^A and U^P in the definition of $\hat{f}$ and G denote extensions from $[\underline{\pi}, \bar{\pi}]$ to all of $\mathbb{R}$ as twice continuously differentiable, strictly concave functions that preserve the single-crossing condition and the relative concavity bound κ. The interval $[\underline{\pi}, \bar{\pi}]$ was chosen so that this extension is without loss.

bounded non-decreasing functions. Let $\mathcal{P} = \{z \in Z \mid z(\theta) \geq 0 \text{ for } \theta \in \Theta\}$ denote the positive cone in the normed space Z with interior point $z(\theta) = 1$ for $\theta \in \Theta$. Let Γ denote the subset of constraints on money burning such that the constraint set is non-empty, $\Gamma = \{z \in Z :$ there exists $(\pi, \underline{\tau}) \in \Omega$ such that $G(\pi, \underline{\tau}) \geq z\}$. The primal functional $\omega : \Gamma \mapsto \mathbb{R}$ is:

$$\omega(z) = \sup\{\hat{f}(\pi, \underline{\tau}) : (\pi, \underline{\tau}) \in \Omega, G(\pi, \underline{\tau}) \geq z\}.$$

The program of interest corresponds to $\omega(0)$. By assumption that (π, τ) is an optimal rule, $\omega(0)$ is finite. The objective is to show that there exists a functional in the dual of Z—the Lagrange multiplier—that supports the graph of ω at $0 \in Z$ and that there is a usable representation of this functional (i.e., amenable to using the tools of calculus).

To obtain a representation of the Lagrange multiplier functional as an integral (as in the Lagrangian (6)), consider a subset of the family of programs so that the Lagrange multiplier functional is in the dual of $Z_c = \{z \mid z : \Theta \mapsto \mathbb{R} \text{ is continuous}\}$. Let $\Omega_c = \{(\pi, \underline{\tau}) \in \Omega :$ π is continuous$\}$ and note that $G : \Omega_c \mapsto Z_c$ since U^A is continuous. Let $\omega_c(z) = \sup\{\hat{f}(\pi, \underline{\tau}) :$ $(\pi, \underline{\tau}) \in \Omega_c, G(\pi, \underline{\tau}) \geq z\}$.

The Lagrange multiplier obtains as a separating hyperplane between the region A_c below the graph of the primal functional ω_c and the cone B_c with origin $(\omega_c(0), 0)$,

$$\begin{aligned} A_c &= \left\{(r, z) \in \mathbb{R} \times Z_c : r \leq \hat{f}(\pi, \underline{\tau}), z \leq G(\pi, \underline{\tau}) \text{ for some } (\pi, \underline{\tau}) \in \Omega_c\right\} \\ B_c &= \{(r, z) \in \mathbb{R} \times Z_c : r \geq \omega_c(0), z \geq 0\}. \end{aligned}$$

A key requirement for the existence of this separating hyperplane is that A_c and B_c are convex. To show that A_c is convex, let $\alpha \in [0, 1]$ and $(r_i, z_i) \in A_c$ for $i = 1, 2$ with $(\pi_i, \underline{\tau}_i) \in \Omega_c$ such that $r_i \leq \hat{f}(\pi_i, \underline{\tau}_i)$, and $z_i(\theta) \leq G(\pi_i, \underline{\tau}_i)$. The next step shows that $(\alpha r_1 + (1-\alpha) r_2, \alpha z_1 + (1-\alpha) z_2) \in A_c$. Let $\pi_\alpha = \alpha \pi_1 + (1 - \alpha)\pi_2$ and

$$\underline{\tau}_\alpha = \alpha \underline{\tau}_1 + (1 - \alpha)\underline{\tau}_2 + U^A(\underline{\theta}, \pi_\alpha(\underline{\theta})) - (\alpha U^A(\underline{\theta}, \pi_1(\underline{\theta})) + (1 - \alpha) U^A(\underline{\theta}, \pi_2(\underline{\theta}))).$$

Hence, $(\pi_\alpha, \underline{\tau}_\alpha) \in \Omega_c$. Because $U^P - \rho U^A$ is concave, and by construction of $\underline{\tau}_\alpha$,

$$\alpha r_1 + (1 - \alpha) r_2 \leq \alpha \hat{f}(\pi_1, \underline{\tau}_1) + (1 - \alpha)\hat{f}(\pi_2, \underline{\tau}_2) \leq \hat{f}(\pi_\alpha, \underline{\tau}_\alpha).$$

Also, the concavity of U^A and the definition of $\underline{\tau}_\alpha$ imply the following inequality

$$\begin{aligned} &\alpha G(\pi_1, \underline{\tau}_1)(\theta) + (1 - \alpha) G(\pi_2, \underline{\tau}_2)(\theta) \\ &\qquad = \underline{\tau}_\alpha - U^A(\underline{\theta}, \pi_\alpha(\underline{\theta})) + \alpha U^A(\theta, \pi_1(\theta)) + (1 - \alpha) U^A(\theta, \pi_2(\theta)) - \int_{\underline{\theta}}^{\theta} \pi_\alpha(\tilde{\theta}) d\tilde{\theta} \\ &\qquad \leq \underline{\tau}_\alpha - U^A(\underline{\theta}, \pi_\alpha(\underline{\theta})) + U^A(\theta, \pi_\alpha(\theta)) - \int_{\underline{\theta}}^{\theta} \pi_\alpha(\tilde{\theta}) d\tilde{\theta} = G(\pi_\alpha, \underline{\tau}_\alpha)(\theta). \end{aligned}$$

It follows that

$$\alpha z_1(\theta) + (1-\alpha) z_2(\theta) \leq \alpha G(\pi_1, \underline{\tau}_1)(\theta) + (1-\alpha) G(\pi_2, \underline{\tau}_2)(\theta) \leq G(\pi_\alpha, \underline{\tau}_\alpha)(\theta),$$

which completes the proof that A_c is convex. The set B_c is convex because Z_c is a convex set.

The convex set A_c contains no interior points of the convex set B_c. Also, the set B_c contains an interior point $(\omega_c(0)+1, z) \in B_c$ where $z(\theta) = 1$ for $\theta \in \Theta$. The geometric Hahn-Banach Theorem (Luenberger (1969) Theorem 1 p. 133) implies that there is a nonzero $(r_0, \Lambda^*) \in \mathbb{R} \times Z_c^*$ that separates the sets A_c and B_c as follows:

$$r_0 r_a + \Lambda^*(z_a) \leq r_0 r_b + \Lambda^*(z_b) \tag{15}$$

for $(r_a, z_a) \in A_c$ and for $(r_b, z_b) \in B_c$.

The definition of B_c is useful to show that $r_0 \geq 0$ and $\Lambda^* \geq 0$. Suppose, by way of contradiction, that $r_0 < 0$. Let $(r_b, z_b) \in B_c$ with $r_b > 0$ and sufficiently large such that it contradicts (15) for some $(r_a, z_a) \in A_c$. Similarly, suppose that Λ^* is not positive, then let $(r_b, z_b) \in B_c$ with z_b such that $\Lambda^*(z_b) < 0$ and sufficiently large that it contradicts (15) for some $(r_a, z_a) \in A_c$.

The existence of an interior point in B_c is useful to show that $r_0 > 0$. For $(\omega_c(0), 0) \in B_c$, inequality (15) implies

$$r_0 r_a + \Lambda^*(z_a) \leq r_0 \omega_c(0).$$

Suppose by way of contradiction that $r_0 = 0$. The point $(\pi_f, 1) \in \Omega_c$ maps into a constant money burning schedule in the interior of $\mathcal{P}$, which is $G(\pi_f, 1)(\theta) = 1$ for $\theta \in \Theta$. Because $(\hat{f}(\pi_f, 1), G(\pi_f, 1)) \in A_c$ and $(\omega(0), 0) \in B_c$, inequality (15) with $r_0 = 0$ implies $\Lambda^*(G(\pi_f, 1)) \leq 0$. However, since $G(\pi_f, 1) > 0$, $\Lambda^* \geq 0$, and $\Lambda^* \neq 0$, it follows that $\Lambda^*(G(\pi_f, 1)) > 0$, which is a contradiction. The constant $r_0 > 0$ can be normalized to $r_0 = 1$ by dividing the functional Λ^* by r_0.

The Lagrange multiplier functional $\Lambda^* \in Z_c^*$ admits an integral representation. Based on the representation of the positive cone in the dual of $C(\Theta)$ for a compact interval Θ, the Riesz representation theorem implies that there exists a non-decreasing function $\Lambda : \Theta \to \mathbb{R}$, right-continuous on $(\underline{\theta}, \bar{\theta})$, such that

$$\Lambda^*(z) = \rho \int_\Theta z(\theta) \mathrm{d}\Lambda(\theta), \tag{16}$$

for $z \in Z_c$, where $\rho > 0$ is a scaling factor (see Theorem 1 p. 113 in Luenberger (1969) for Riesz representation and p. 215 in Luenberger (1969) for the positive cone in the dual of $C(\Theta)$). We will show later that the scaling factor normalizes $\Lambda(\bar{\theta}) = 1$.

The functional Λ^* represented by the non-decreasing function Λ in (16) is well-defined for any $z \in Z$ because the Lebesgue-Stieltjes integral (16) is well-defined for z of bounded variation.

The point $(\omega_c(0), 0)$ is at the boundary of A_c and B_c, hence

$$r_a + \rho \int_\Theta z_a(\theta) \mathrm{d}\Lambda(\theta) \leq \omega_c(0) \leq r_b + \rho \int_\Theta z_b(\theta) \mathrm{d}\Lambda(\theta) \tag{17}$$

for $(r_a, z_a) \in A_c$ and for $(r_b, z_b) \in B_c$.

An approximation argument shows that the Lagrange multiplier functional (16) on Z is a supporting hyperplane of

$$A = \left\{ (r, z) \in \mathbb{R} \times Z : r \leq \hat{f}(\pi, \underline{\tau}), z \leq G(\pi, \underline{\tau}) \text{ for some } (\pi, \underline{\tau}) \in \Omega \right\}.$$

Suppose, by way of contradiction, that there exists $(r, z) \in A$ such that

$$r + \rho \int_\Theta z(\theta) \mathrm{d}\Lambda(\theta) > \omega_c(0),$$

and let $\epsilon = r + \rho \int_\Theta z(\theta) \mathrm{d}\Lambda(\theta) - \omega_c(0) > 0$. The approximation argument constructs an element of A_c that also satisfies the strict inequality, which is a contradiction to (17).

Intuitively, for $(r, z) \in A$, by definition of A, there exists $(\pi, \underline{\tau}) \in \Omega$ such that $r \leq \hat{f}(\pi, \underline{\tau})$ and $z \leq G(\pi, \underline{\tau})$. The argument shows that there exists $(r_c, z_c) \in A_c$ such that $|\Lambda^*(z) - \Lambda^*(z_c)| < \frac{\epsilon}{2}$ and $|r - r_c| < \frac{\epsilon}{2}$. To do so, in the spirit of Lusin's theorem, I show that there exists an admissible rule with a continuous policy function π_c that agrees with π on a set of measure arbitrarily close to full measure. Intuitively, π_c ramps up over small intervals around points of discontinuity of π and coincides with π otherwise. Because π has at most countably many points of discontinuity, it is possible to choose the intervals over which π_c ramps up to be small enough that they have arbitrarily small mass under the measures generated by the distributions F and Λ. In particular, the construction of π_c is such that $|\hat{f}(\pi_c, \underline{\tau}) - \hat{f}(\pi, \underline{\tau})| < \epsilon/2$ and $|\Lambda^*(G(\pi_c, \underline{\tau})) - \Lambda^*(G(\pi, \underline{\tau}))| < \epsilon/2$. This contradicts inequality (17) for $(\hat{f}(\pi_c, \underline{\tau}), G(\pi_c, \underline{\tau})) \in A_c$.

Formally, let $\{\theta_i\}_{i \in I}$, denote the set of points of discontinuity of π. The set I is countable because π is non-decreasing on $[\underline{\theta}, \bar{\theta}]$. To build a continuous function π_c that approximates π, construct the ramp intervals sequentially, following an enumeration of I. If θ_i already lies in a previously constructed interval for some $j < i$, no separate ramp is needed at θ_i (π_c is already continuous there). Relabel the remaining index set accordingly and proceed to the next i. Otherwise, for a point at which π is left-continuous, having fixed $w_1, \ldots, w_{i-1}$, let $w_i > 0$ be small enough that $\theta_i + w_i \leq \bar{\theta}$, $\theta_i + w_i$ is a point of continuity of π, and $(\theta_i, \theta_i + w_i]$ does not intersect any of the finitely many previously constructed intervals $j < i$ (possible because θ_i is distinct from, and lies outside the finitely many intervals around, each of the finitely many θ_j, $j < i$). To construct the ramp that continuously goes from $\pi(\theta_i)$ to $\pi(\theta_i + w_i)$ over the interval $(\theta_i, \theta_i + w_i]$, let $h_i = \pi(\theta_i + w_i) - \pi(\theta_i) \geq 0$ denote an upper bound for the jump at θ_i (h_i is finite

because π is bounded, and $h_i \geq 0$ because π is non-decreasing). The key is to choose the width $w_i > 0$ small enough such that the resulting π_c is close to π in terms of welfare for the principal, and to place the ramp so that $\Lambda^*(G(\pi_c, \underline{\tau}))$ is close to $\Lambda^*(G(\pi, \underline{\tau}))$. Let

$$M_P = 2 \max_{(\theta,\pi)\in[\underline{\theta},\bar{\theta}]\times[\underline{\pi},\bar{\pi}]} \left\{ \left| U^P(\theta,\pi) - \rho U^A(\theta,\pi) + \rho\pi \frac{1-F(\theta)}{f(\theta)} \right| \right\}.$$

and

$$M_A = 2 \max_{(\theta,\pi)\in[\underline{\theta},\bar{\theta}]\times[\underline{\pi},\bar{\pi}]} \big\{ |U^A(\theta,\pi)| \big\}.$$

Let μ_Λ denote the finite measure generated by the non-decreasing right-continuous function Λ. Note that $\lim_{w_i\to 0} \mu_\Lambda\big((\theta_i, \theta_i + w_i]\big) = 0$ because

$$\mu_\Lambda\big((\theta_i, \theta_i + w_i]\big) = \int_{(\theta_i,\theta_i+w_i]} 1 d\Lambda(\theta) = \Lambda(\theta_i + w_i) - \Lambda(\theta_i^+).$$

Choose w_i small enough such that $F(\theta_i + w_i) - F(\theta_i) < \frac{\epsilon}{2^{i+1} M_P}$,

$$\mu_\Lambda\big((\theta_i, \theta_i + w_i]\big) < \frac{\epsilon}{2^{i+2} M_A} \quad \text{and} \quad w_i < \frac{\epsilon}{2^{i+2} h_i \max(1, (\Lambda(\bar{\theta}) - \Lambda(\underline{\theta})))}.$$

By construction, the intervals $(\theta_i, \theta_i + w_i]$, $i \in I$, are pairwise disjoint and contained in $[\underline{\theta}, \bar{\theta}]$. Let $\Theta_c = [\underline{\theta}, \bar{\theta}] \setminus \cup_{i\in I} (\theta_i, \theta_i + w_i]$ and define

$$\pi_c(\theta) = \begin{cases} \pi(\theta) & \text{for } \theta \in \Theta_c \\ \pi(\theta_i) + \frac{\theta - \theta_i}{w_i} (\pi(\theta_i + w_i) - \pi(\theta_i)) & \text{for } \theta \in (\theta_i, \theta_i + w_i]. \end{cases}$$

By construction, π_c and π achieve welfare for the principal that is at most $\frac{\epsilon}{2}$ apart,

$$\begin{aligned} |\hat{f}(\pi_c, \underline{\tau}) - \hat{f}(\pi, \underline{\tau})| &\leq \sum_{i\in I} \int_{(\theta_i,\theta_i+w_i)} \Big| U^P(\theta, \pi_c(\theta)) - U^P(\theta, \pi(\theta)) - \rho(U^A(\theta, \pi_c(\theta)) - U^A(\theta, \pi(\theta))) \\ &\qquad + \rho(\pi_c(\theta) - \pi(\theta)) \frac{1-F(\theta)}{f(\theta)} \Big| dF(\theta) \\ &< \sum_{i\in I} M_P \frac{\epsilon}{2^{i+1} M_P} \leq \frac{\epsilon}{2}. \end{aligned}$$

Similarly,

$$\begin{aligned} |\Lambda^*(G(\pi_c, \underline{\tau})) - \Lambda^*(G(\pi, \underline{\tau}))| &\leq \rho \sum_{i\in I} \int_{(\theta_i,\theta_i+w_i)} |U^A(\theta, \pi_c(\theta)) - U^A(\theta, \pi(\theta))| d\Lambda(\theta) \\ &\quad + \rho(\Lambda(\bar{\theta}) - \Lambda(\underline{\theta})) \sum_{i\in I} \int_{(\theta_i,\theta_i+w_i)} |\pi_c(\tilde{\theta}) - \pi(\tilde{\theta})| d(\tilde{\theta}) \\ &< \rho \sum_{i\in I} M_A \frac{\epsilon}{2^{i+2} M_A} + \rho \sum_{i\in I} h_i \frac{\epsilon}{2^{i+2} h_i} \leq \rho \frac{\epsilon}{2}. \end{aligned}$$

The construction of the ramp at points at which π is right-continuous is symmetric.

It follows that

$$\hat{f}(\pi_c, \underline{\tau}) + \Lambda^*(G(\pi_c, \underline{\tau})) > \hat{f}(\pi, \underline{\tau}) + \Lambda^*(G(\pi, \underline{\tau})) - \epsilon \geq \omega_c(0).$$

However, this contradicts (17) because $(\hat{f}(\pi_c, \underline{\tau}), G(\pi_c, \underline{\tau})) \in A_c$ since π_c is continuous and non-decreasing. This concludes the argument that for $(r_a, z_a) \in A$,

$$r_a + \rho \int_\Theta z_a(\theta) \mathrm{d}\Lambda(\theta) \leq \omega_c(0). \tag{18}$$

The next step shows that the solution maximizes the Lagrangian; that is

$$\omega(0) = \sup_{(\pi, \underline{\tau}) \in \Omega} \left\{ \hat{f}(\pi, \underline{\tau}) + \rho \int_\Theta G(\pi, \underline{\tau})(\theta) \mathrm{d}\Lambda(\theta) \right\}. \tag{19}$$

To see this, note that $\omega(0) \geq \omega_c(0)$ and

$$\begin{aligned}
\omega_c(0) &\geq \sup_{(r_a, z_a) \in A} \left\{ r_a + \rho \int_\Theta z_a(\theta) \mathrm{d}\Lambda(\theta) \right\} \\
&\geq \sup_{(\pi, \underline{\tau}) \in \Omega} \left\{ \hat{f}(\pi, \underline{\tau}) + \rho \int_\Theta G(\pi, \underline{\tau})(\theta) \mathrm{d}\Lambda(\theta) \right\} \\
&\geq \sup_{(\pi, \underline{\tau}) \in \Omega} \left\{ \hat{f}(\pi, \underline{\tau}) + \rho \int_\Theta G(\pi, \underline{\tau})(\theta) \mathrm{d}\Lambda(\theta) : G(\pi, \underline{\tau}) \in \mathcal{P} \right\}, \\
&\geq \sup_{(\pi, \underline{\tau}) \in \Omega} \left\{ \hat{f}(\pi, \underline{\tau}) : G(\pi, \underline{\tau}) \in \mathcal{P} \right\} = \omega(0),
\end{aligned}$$

where the first inequality follows from (18), and the second inequality follows from the definition of A, which implies that $(\hat{f}(\pi, \underline{\tau}), G(\pi, \underline{\tau})) \in A$ for $(\pi, \underline{\tau}) \in \Omega$. The third inequality holds because adding a constraint can only decrease the value. The fourth inequality follows from the previous result that Λ is non-decreasing and $G(\pi, \underline{\tau}) \in \mathcal{P}$. The last equality holds by definition of the primal functional ω.

The complementary slackness condition (iii) holds because $(\pi, \underline{\tau})$ achieves the supremum (by assumption) in the definition of $\omega(0)$ and in (19), which implies that $\int_\Theta G(\pi, \underline{\tau})(\theta) \mathrm{d}\Lambda(\theta) = 0$.

To show that $\Lambda(\underline{\theta}) \geq 0$, I show that $\Lambda(\underline{\theta}) = 0$ because Λ is non-decreasing. Suppose, by way of contradiction, that $\rho\Lambda(\underline{\theta}) \neq 0$ (if $\rho = 0$, the proof follows from first principles). Let $|h_\tau|$ be sufficiently large such that it contradicts $\partial\mathcal{L}(\pi, \tau(\underline{\theta}), h, h_\tau) \leq 0$ based on (7).

To show that $\Lambda(\bar{\theta}) = 1$, I use the sensitivity of the primal functional to a tightening of the money-burning constraint. Consider a uniform tightening of the money-burning constraint, denoted $1_{[\underline{\theta}, \bar{\theta}]}(\theta) = 1$ for $\theta \in [\underline{\theta}, \bar{\theta}]$. The argument shows that the primal functional is Gateaux differentiable at $0 \in Z$ in the direction $1_{[\underline{\theta}, \bar{\theta}]}$. This is because a uniform tightening of the money-burning constraint simply shifts the entire money-burning schedule by a constant. Formally,

let $(\pi, \underline{\tau})$ be such that it achieves the supremum in the definition of ω for the original program $z(\theta) = 0$ for $\theta \in [\underline{\theta}, \bar{\theta}]$ and let Λ be the associated multiplier function. By inspection, $(\pi, \underline{\tau}+1)$ also achieves the supremum in the definition of ω for the program with the tightened money-burning constraint $1_{[\underline{\theta},\bar{\theta}]}$. By direct computation $\omega(\alpha 1_{[\underline{\theta},\bar{\theta}]}) = \omega(0) - \alpha\rho$ and

$$\partial\omega(0, 1_{[\underline{\theta},\bar{\theta}]}) = \lim_{\alpha \to 0} \frac{1}{\alpha}[\omega(\alpha 1_{[\underline{\theta},\bar{\theta}]}) - \omega(0)] = -\rho.$$

The supporting hyperplane Λ^* is a supergradient of the primal functional $\omega(z) \leq \omega(0) - \Lambda^*(z)$. If the primal functional is differentiable in the direction z, the supporting hyperplane Λ^* is a gradient of the primal functional

$$\partial\omega(0, z) = -\Lambda^*(z) = -\rho \int_{\underline{\theta}}^{\bar{\theta}} z(\theta) d\Lambda(\theta).$$

In particular, because the primal functional is differentiable in the direction $1_{[\underline{\theta},\bar{\theta}]}$,

$$\partial\omega(0, 1_{[\underline{\theta},\bar{\theta}]}) = -\rho(\Lambda(\bar{\theta}) - \Lambda(\underline{\theta})) = -\rho\Lambda(\bar{\theta}),$$

and hence $\Lambda(\bar{\theta}) = 1$.

The second step derives necessary optimality conditions (i) and (ii) given that $(\pi, \tau(\underline{\theta}))$ maximizes the Lagrangian with the Riesz representation of the Lagrange multiplier functional (16). For $(h, h_\tau) \in \Omega$ and $\alpha \geq 0$, $(\pi + \alpha h, \tau(\underline{\theta}) + \alpha h_\tau) \in \Omega$. Because $(\pi, \tau(\underline{\theta}))$ maximizes the Lagrangian over Ω, condition (i) necessarily holds; that is, $\partial\mathcal{L}(\pi, \tau(\underline{\theta}), h, h_\tau) \leq 0$ for $(h, h_\tau) \in \Omega$.

Also, $\partial\mathcal{L}(\pi, \tau(\underline{\theta}), -\pi, -\tau(\underline{\theta})) \leq 0$ because $(\pi - \alpha\pi, \tau(\underline{\theta}) - \alpha\tau(\underline{\theta})) \in \Omega$ for $\alpha \in [0, 1]$. Because the Gateaux derivative is linear in its increment,

$$-\partial\mathcal{L}(\pi, \tau(\underline{\theta}), \pi, \tau(\underline{\theta})) = \partial\mathcal{L}(\pi, \tau(\underline{\theta}), -\pi, -\tau(\underline{\theta})) \leq 0.$$

Also, $\partial\mathcal{L}(\pi, \tau(\underline{\theta}), \pi, \tau(\underline{\theta})) \leq 0$ because $(\pi + \alpha\pi, \tau(\underline{\theta}) + \alpha\tau(\underline{\theta})) \in \Omega$. As a result, condition (ii) holds.

Condition (iv) holds because an optimal rule is, by definition, admissible.

A.1.2 Sufficiency

The proof proceeds in three steps. The first step verifies that the Lagrangian is concave. The second step shows that the optimality conditions imply that the Lagrangian is maximized. The third step shows that, if the complementary slackness condition is satisfied, the maximizer of a Lagrangian is the solution to the constrained optimization problem of interest.

First, to show the concavity of the Lagrangian given any non-decreasing Lagrange multiplier Λ, it is convenient to substitute the money burning schedule (4) in the Lagrangian (6) and factorize the non-linear terms as follows:

$$\begin{aligned}\mathcal{L}(\pi,\underline{\tau}|\Lambda) = &\int_{\Theta}\big[U^P(\theta,\pi(\theta)) - \kappa U^A(\theta,\pi(\theta))\big]dF(\theta)\\ &+\rho\int_{\Theta}\big[\pi(\theta)(\Lambda(\theta)-F(\theta))\big]d\theta\\ &+\int_{\Theta}U^A(\theta,\pi(\theta))d\big((\kappa-\rho)F(\theta)+\rho\Lambda(\theta)\big)\\ &+\rho\left[U^A(\underline{\theta},\pi(\underline{\theta}))-\underline{\tau}\right]\Lambda(\underline{\theta}).\end{aligned}$$

so that the integrand of the first integral in the Lagrangian is concave by definition of κ (following Amador and Bagwell (2013)). The integrand of the second integral is concave because it is linear in π. The integrand of the third integral is concave because U^A is concave, F and Λ are both non-decreasing and, by assumption, $\kappa \geq \rho$.

Second, Lemma A.2 in Amador, Werning, and Angeletos (2006) implies that if the Lagrangian is concave (given the Lagrange multiplier Λ), then the conditions (i) and (ii) imply that (π,τ) maximizes the Lagrangian $\mathcal{L}(\pi,\underline{\tau}|\Lambda) \geq \mathcal{L}(\tilde{\pi},\tilde{\underline{\tau}}|\Lambda)$ for $(\tilde{\pi},\tilde{\underline{\tau}}) \in \Omega$.

Third, the global theory of constrained optimization implies that, if the complementary slackness condition is satisfied, the maximizer of a Lagrangian at a valid Lagrange multiplier is the solution to the associated constrained optimization problem. To apply Theorem 1 in Amador and Bagwell (2013) p. 1575, let $T : Z \mapsto \mathbb{R}$,

$$T(z) = \rho\int_{\Theta} z(\theta)d\Lambda(\theta).$$

Note that $T(z) \geq 0$ for all $z \in \mathcal{P}$ because Λ is non-decreasing. The Lagrangian read $\mathcal{L}(\pi,\underline{\tau}|\Lambda) = \hat{f}(\pi,\underline{\tau}) + T(G(\pi,\underline{\tau}))$, and, as shown above, $\mathcal{L}(\pi,\underline{\tau}|\Lambda) \geq \mathcal{L}(\tilde{\pi},\tilde{\underline{\tau}}|\Lambda)$ for $(\tilde{\pi},\tilde{\underline{\tau}}) \in \Omega$. Hence, $\hat{f}(\pi,\underline{\tau}) + T(G(\pi,\underline{\tau})) \geq \hat{f}(\tilde{\pi},\tilde{\underline{\tau}}) + T(G(\tilde{\pi},\tilde{\underline{\tau}}))$ for $(\tilde{\pi},\tilde{\underline{\tau}}) \in \Omega$. The complementary slackness condition holds by condition (iii), and $G(\pi,\tau(\underline{\theta})) \in \mathcal{P}$ because (π,τ) is admissible by condition (iv). Hence, Theorem 1 from Amador and Bagwell (2013) implies that $(\pi,\underline{\tau})$ solves

$$\max_{(\tilde{\pi},\tilde{\underline{\tau}})\in\Omega}\{\hat{f}(\tilde{\pi},\tilde{\underline{\tau}})|G(\tilde{\pi},\tilde{\underline{\tau}}) \in \mathcal{P}\}.$$

That is, (π,τ) is an optimal rule. □

A.2 Proof of Lemma 2

A.2.1 Necessity

The proof that the conditions in Lemma 2 are necessary is in three parts. The first part of the proof shows that the condition (i) in Lemma 1 implies the condition (i) in Lemma 2.

To show that $\delta(\theta) \leq 0$ for $\theta \in (\underline{\theta}, \bar{\theta})$, let h be a step function with $h(\underline{\theta}) = 0$ and one positive left-continuous jump at $\theta \in (\underline{\theta}, \bar{\theta})$ (the jump is left-continuous so that $d(\theta) = 0$ because δ is right-continuous). The Gateaux derivative of the Lagrangian (9) simplifies to

$$\partial\mathcal{L}(\pi, \underline{\tau}, h, h_\tau) = \delta(\theta)(h(\theta^+) - h(\theta^-)).$$

Condition (i) in Lemma 1 implies that $\delta(\theta) \leq 0$ for $\theta \in (\underline{\theta}, \bar{\theta})$ because h is non-decreasing.

To show that $\delta(\underline{\theta}) = \delta(\bar{\theta})$, consider a constant perturbation $h(\theta) = 1$ for $\theta \in \Theta$ and $h_\tau = 0$. Condition (i) in Lemma 1 implies

$$\delta(\underline{\theta}) - \delta(\bar{\theta}) = \partial\mathcal{L}(\pi, \underline{\tau}, h, h_\tau) \leq 0.$$

Similarly, $h(\theta) = -1$ for $\theta \in \Theta$ and $h_\tau = 0$ implies that $\delta(\underline{\theta}) - \delta(\bar{\theta}) \geq 0$. Hence, $\delta(\underline{\theta}) = \delta(\bar{\theta}) = 0$.

Condition (ii) in Lemma 1 implies that

$$0 = \int_{\underline{\theta}}^{\bar{\theta}} \delta(\theta) d\pi(\theta) - \sum_{\theta \in D_{\pi\delta}} d(\theta). \tag{20}$$

The argument is that the contribution of each of the two terms on the right-hand side of (20) is negative, hence each term has to be equal to zero. To see this, let D_π denote the set of points of discontinuity of π, hence

$$\int_{\underline{\theta}}^{\bar{\theta}} \delta(\theta) d\pi(\theta) = \int_{\Theta \setminus D_\pi} \delta(\theta) d\pi(\theta) + \sum_{\theta \in D_\pi} \left[\delta(\theta)(\pi(\theta^+) - \pi(\theta^-))\right]. \tag{21}$$

Note that, because δ is right-continuous on $(\underline{\theta}, \bar{\theta})$,

$$\begin{aligned} &\sum_{\theta \in D_\pi} [\delta(\theta)(\pi(\theta^+) - \pi(\theta^-))] - \sum_{\theta \in D_{\pi\delta}} d(\theta) \\ &= \sum_{\theta \in D_\pi \setminus D_{\pi\delta}} [\delta(\theta)(\pi(\theta^+) - \pi(\theta^-))] + \sum_{\theta \in D_{\pi\delta}} \left[\delta(\theta^-)(\pi(\theta) - \pi(\theta^-)) + \delta(\theta)(\pi(\theta^+) - \pi(\theta))\right]. \end{aligned} \tag{22}$$

Hence, because $\delta(\theta) \leq 0$ for $\theta \in \Theta$, π is non-decreasing, and $\delta(\underline{\theta}) = 0$, the right-hand side of (20), decomposed as in (21) and (22), is a sum of non-positive terms equal to zero; hence each term is zero. In particular,

$$\int_{\Theta \setminus D_\pi} \delta(\theta) d\pi(\theta) = 0, \tag{23}$$

and, for each $\theta \in D_{\pi\delta}$, $\delta(\theta^-)(\pi(\theta) - \pi(\theta^-)) + \delta(\theta)(\pi(\theta^+) - \pi(\theta)) = 0$.

It remains to show that $D_{\pi\delta} = \emptyset$, so that δ is continuous at every point of discontinuity of π. Suppose, by way of contradiction, that $\theta \in D_{\pi\delta}$. If $\pi(\theta) \in (\pi(\theta^-), \pi(\theta^+))$, both increments in the previous equality are positive, so $\delta(\theta^-) = \delta(\theta) = 0$ and δ is continuous at θ, a contradiction. Hence $\pi(\theta)$ is an endpoint of the jump; suppose $\pi(\theta) = \pi(\theta^-)$ (the case $\pi(\theta) = \pi(\theta^+)$ is symmetric), so that the equality gives $\delta(\theta) = 0$. Because δ is discontinuous at θ, the identity $\delta(\theta^-) - \delta(\theta) = \rho\Delta(\theta, \pi(\theta))(\Lambda(\theta) - \Lambda(\theta^-))$, which follows from the definition of δ, implies that Λ jumps at θ; the complementary slackness condition (iii) in Lemma 1 then gives $\tau(\theta) = 0$. By the envelope condition (4), $\tau(\theta^+) = U^A(\theta, \pi(\theta^+)) - U^A(\theta, \pi(\theta^-))$, which is non-negative because $\tau \geq 0$; by strict concavity of $U^A(\theta, \cdot)$ and $\pi(\theta^-) < \pi(\theta^+)$, this implies $\pi(\theta^-) < \pi_f(\theta)$, hence $\Delta(\theta, \pi(\theta)) > 0$. But then $\delta(\theta^-) = \delta(\theta) + \rho\Delta(\theta, \pi(\theta))(\Lambda(\theta) - \Lambda(\theta^-)) > 0$, contradicting that $\delta(\theta^-) \leq 0$ as the left limit of $\delta \leq 0$. Therefore $D_{\pi\delta} = \emptyset$: the sum (22) reduces to $\sum_{\theta \in D_\pi} \delta(\theta)(\pi(\theta^+) - \pi(\theta^-)) = 0$, and $\delta(\theta^-) = \delta(\theta) = 0$ at any point of discontinuity of π.

For any interval $[\theta_1, \theta_2) \subseteq \Theta$ over which π is strictly increasing, condition (23) implies that $\delta(\theta) = 0$ for $\theta \in [\theta_1, \theta_2)$. At the countably many discontinuities of π in $[\theta_1, \theta_2)$, $\delta(\theta) = 0$ as shown above. Suppose, by way of contradiction, that $\delta(\theta) < 0$ for some point of continuity $\theta \in [\theta_1, \theta_2)$. Since π is strictly increasing over $[\theta_1, \theta_2)$ and δ is right-continuous on $(\underline{\theta}, \bar{\theta})$, $\int_{[\theta_1, \theta_2) \setminus D_\pi} \delta(\theta) d\pi(\theta) < 0$. However, this contradicts (23) because $\delta(\theta) \leq 0$ for $\theta \in \Theta$ and π is non-decreasing.

A.2.2 Sufficiency

Condition (i) in Lemma 2 and $\Lambda(\underline{\theta})$ imply that the Gateaux derivative of the Lagrangian (9) simplifies to

$$\partial\mathcal{L}(\pi, \tau(\underline{\theta}), h, h_\tau) = \int_{(\underline{\theta}, \bar{\theta}]} \delta(\theta) dh(\theta) - \sum_{\theta \in D_{h\delta}} d(\theta). \tag{24}$$

To show that condition (i) in Lemma 1 holds, note that the integral is non-positive because h is non-decreasing and $\delta(\theta) \leq 0$ because condition (i) in Lemma 2 holds. At any common point of discontinuity θ of h and δ, the contribution of θ to the right-hand side of (24) is $\int_{[\theta, \theta]} \delta(\theta)\, dh(\theta) - d(\theta)$. Because δ is right-continuous, the definition of $d(\theta)$ simplifies to $d(\theta) = (\delta(\theta) - \delta(\theta^-))(h(\theta) - h(\theta^-))$, so the contribution equals

$$\delta(\theta)\left(h(\theta^+) - h(\theta)\right) + \delta(\theta^-)\left(h(\theta) - h(\theta^-)\right) \leq 0,$$

since $\delta(\theta) \leq 0$, $\delta(\theta^-) \leq 0$, and h is non-decreasing (so $h(\theta^+) \geq h(\theta) \geq h(\theta^-)$).

Condition (ii) in Lemma 1 holds because condition (ii) in Lemma 2 implies that either $\delta(\theta) = 0$ or π is constant, there are no common points of discontinuity of δ and π. □

A.3 Proof of Proposition 1

The complementary slackness condition implies that there exists a constant $\lambda \in [0,1]$ such that $\Lambda(\theta) = \lambda$ for $\theta \in (\theta_1, \theta_2)$, because $\tau(\theta) > 0$ on (θ_1, θ_2) and Λ is non-decreasing.

The optimality condition (ii) in Lemma 2 implies that $\delta(\theta) = 0$ for $\theta \in (\theta_1, \theta_2)$ for a non-decreasing Lagrange multiplier function Λ. By linearity of the integral,

$$\int_{[\hat{\theta},\theta_2)} \left[U_\pi^P(\theta, \pi(\theta)) - \rho\Delta(\theta, \pi(\theta)) + \rho \frac{\Lambda(\theta) - F(\theta)}{f(\theta)} \right] dF(\theta) + \rho \int_{[\hat{\theta},\theta_2)} \Delta(\theta, \pi(\theta)) d\Lambda(\theta) = 0$$

for $\hat{\theta} \in (\theta_1, \theta_2)$. The constant Lagrange multiplier implies that the condition reduces to

$$\int_{[\hat{\theta},\theta_2)} \left[U_\pi^P(\theta, \pi(\theta)) - \rho\Delta(\theta, \pi(\theta)) + \rho \frac{\lambda - F(\theta)}{f(\theta)} \right] dF(\theta) = 0$$

for $\hat{\theta} \in (\theta_1, \theta_2)$. It follows that condition (10), which using $U_\pi^P = \Delta - \nu_\pi$ can be rewritten as

$$\rho\Delta(\theta, \pi(\theta)) = U_\pi^P(\theta, \pi(\theta)) + \rho \frac{\lambda - F(\theta)}{f(\theta)}$$

holds almost everywhere on (θ_1, θ_2). In turn, condition (10) holds everywhere on (θ_1, θ_2) because U_π^P, Δ, F, and f are continuous, $U_\pi^P - \rho\Delta$ is non-increasing in π (since $\rho \leq \kappa$), and π is non-decreasing.

Condition (11) obtains from the monotonicity of π. Let

$$G(\theta, \pi) = U_\pi^P(\theta, \pi) - \rho\Delta(\theta, \pi) + \rho \frac{\lambda - F(\theta)}{f(\theta)},$$

so that (10) reads $G(\theta, \pi(\theta)) = 0$ for $\theta \in (\theta_1, \theta_2)$, and G is non-increasing in π because $G_\pi = U_{\pi\pi}^P - \rho U_{\pi\pi}^A \leq 0$ (since $\rho \leq \kappa$). Fix $\theta \in (\theta_1, \theta_2)$ and $\theta' \in (\theta, \theta_2)$. Because π is non-decreasing, $\pi(\theta') \geq \pi(\theta)$, so

$$G(\theta', \pi(\theta)) \geq G(\theta', \pi(\theta')) = 0 = G(\theta, \pi(\theta)).$$

Dividing $G(\theta', \pi(\theta)) - G(\theta, \pi(\theta)) \geq 0$ by $\theta' - \theta > 0$ and letting $\theta' \downarrow \theta$ gives $G_\theta(\theta, \pi(\theta)) \geq 0$, where the partial derivative is taken at fixed second argument $\pi(\theta)$ and exists because U_π^P, F, and f are continuously differentiable in θ. Since $\Delta_\theta = U_{\pi\theta}^A = 1$,

$$G_\theta(\theta, \pi(\theta)) = U_{\theta\pi}^P(\theta, \pi(\theta)) - \rho + \rho \frac{d}{d\theta} \frac{\lambda - F(\theta)}{f(\theta)},$$

so $G_\theta(\theta, \pi(\theta)) \geq 0$ is inequality (11). If G is independent of π over a subinterval of (θ_1, θ_2), then $G(\theta', \pi(\theta)) = G(\theta', \pi(\theta')) = 0$, and the same argument gives $G_\theta(\theta, \pi(\theta)) = 0$, so (11) holds as an equality.

The complementary slackness condition determines the value of λ in some cases. If $\tau(\theta) > 0$ for $\theta \in (\theta_1, \bar{\theta}]$, then $\lambda = \Lambda(\bar{\theta}) = 1$ because the complementary slackness condition implies that Λ is constant over any interval over which $\tau(\theta) > 0$. Similarly, if $\tau(\theta) > 0$ for $\theta \in [\underline{\theta}, \theta_2)$, then $\lambda = \Lambda(\underline{\theta}) = 0$. □

A.4 Proof of Proposition 2

Lemmas 1 and 2 imply that there exists a non-decreasing function $\Lambda : \Theta \to [0,1]$ with $\Lambda(\underline{\theta}) = 0$ and $\Lambda(\bar{\theta}) = 1$ such that $\delta(\theta) \leq 0$ for $\theta \in (\underline{\theta}, \bar{\theta})$, and $\delta(\underline{\theta}) = \delta(\bar{\theta}) = 0$. In addition, the complementary slackness condition (iii) in Lemma 1 holds $\int_\Theta \tau(\theta) \mathrm{d}\Lambda(\theta) = 0$.

The first part of the proof shows that the condition $\delta(\theta) \leq 0$ for $\theta \in [\theta_1, \theta_2]$ can be rewritten as inequality (12). If $\theta_2 < \bar{\theta}$, it is convenient to rewrite $\delta(\theta)$ to focus on the bunching region over $[\theta_1, \theta_2]$. Note that condition (ii) in Lemma 2 implies that $\delta(\theta_2) = 0$. Either π jumps up at θ_2, in which case condition (ii) gives $\delta(\theta_2) = 0$ directly, or π is strictly increasing over $(\theta_2, \theta_2 + \epsilon)$ for some $\epsilon > 0$, in which case condition (ii) gives $\delta(\theta_2) = 0$ by right-continuity of δ. As a result, by linearity of the integral,

$$\begin{aligned}\delta(\theta) = \delta(\theta) - \delta(\theta_2) = &\int_{(\theta,\theta_2]} \left[U_\pi^P(\hat{\theta}, \pi^*) - \rho \Delta(\hat{\theta}, \pi^*) + \rho \frac{1 - F(\hat{\theta})}{f(\hat{\theta})} \right] dF(\hat{\theta}) - \rho \int_{(\theta,\theta_2]} [1 - \Lambda(\hat{\theta})] d\hat{\theta} \\ &+ \rho \int_{(\theta,\theta_2]} \Delta(\hat{\theta}, \pi^*) d\Lambda(\hat{\theta}) + \mathbb{1}(\theta = \underline{\theta}) \rho \Delta(\underline{\theta}, \pi(\underline{\theta}))(\Lambda(\underline{\theta}^+) - \Lambda(\underline{\theta})).\end{aligned}$$

For $\theta \neq \underline{\theta}$, the indicator term vanishes, and the formula for δ simplifies to

$$\begin{aligned}\delta(\theta) = &\int_{(\theta,\theta_2]} \left[U_\pi^P(\hat{\theta}, \pi^*) - \rho \Delta(\hat{\theta}, \pi^*) + \rho \frac{1 - F(\hat{\theta})}{f(\hat{\theta})} \right] dF(\hat{\theta}) \\ &- \rho \int_{(\theta,\theta_2]} [1 - \Lambda(\hat{\theta})] d\hat{\theta} + \rho \int_{(\theta,\theta_2]} \Delta(\hat{\theta}, \pi^*) d\Lambda(\hat{\theta}).\end{aligned}$$

Integrating by parts

$$- \int_{(\theta,\theta_2]} [1 - \Lambda(\hat{\theta})] d\hat{\theta} + \int_{(\theta,\theta_2]} \Delta(\hat{\theta}, \pi^*) d\Lambda(\hat{\theta}) = \Delta(\theta_2, \pi^*)(\Lambda(\theta_2) - 1) - \Delta(\theta, \pi^*)(\Lambda(\theta) - 1).$$

Substituting $U_\pi^P = \Delta - \nu_p$ and $\Delta(\theta_2, \pi^*) = \theta_2 - \theta + \Delta(\theta, \pi^*)$ into $\delta(\theta)$ gives

$$\delta(\theta) = \int_\theta^{\theta_2} \left[(1 - \rho) \Delta(\hat{\theta}, \pi^*) - \nu_\pi^P(\hat{\theta}, \pi^*) + \rho \frac{\Lambda(\theta_2) - F(\hat{\theta})}{f(\hat{\theta})} \right] dF(\hat{\theta}) + \rho \Delta(\theta, \pi^*)(\Lambda(\theta_2) - \Lambda(\theta)). \quad (25)$$

The definition of the step function λ is such that the condition $\delta(\theta) \leq 0$ for $\theta \in [\theta_1, \theta_2]$ based on (25) implies inequality (12) for $\theta \in [\theta_1, \theta_2]$. To see this, let $\underline{\lambda} = \Lambda(\theta_1)$ and $\bar{\lambda} = \Lambda(\theta_2)$. Note that $\Lambda(\theta) \in [\underline{\lambda}, \bar{\lambda}]$ for $\theta \in [\theta_1, \theta_2]$ because Λ is non-decreasing. It follows that $\lambda(\theta) = \underline{\lambda} \leq \Lambda(\theta)$ if $\Delta(\theta, \pi^*) < 0$ and $\lambda(\theta) = \bar{\lambda} \geq \Lambda(\theta)$ if $\Delta(\theta, \pi^*) \geq 0$. As a result, substituting λ for Λ in the formula (25) for $\delta(\theta) \leq 0$ implies inequality (12).

If $\tau(\theta) > 0$ for $\theta \in [\theta_1, \theta_2]$, then the complementary slackness condition implies that Λ is constant over $[\theta_1, \theta_2]$, hence $\Lambda(\theta_1) = \Lambda(\theta_2)$ and $\underline{\lambda} = \bar{\lambda}$.

If $\tau(\theta) > 0$ for $\theta \in (\theta_1, \theta_2)$, then the complementary slackness condition implies that Λ is constant over (θ_1, θ_2), hence $\Lambda(\theta_1^+) = \Lambda(\theta_2^-)$ and $\int_{(\theta_1,\theta_2)} \tau(\theta) \mathrm{d}\Lambda(\theta) = 0$. At the left endpoint, right-continuity of Λ (for $\theta_1 > \underline{\theta}$) gives $\Lambda(\theta_1) = \Lambda(\theta_1^+)$. At the right endpoint, the measure $\mathrm{d}\Lambda$ at θ_2 is $\Lambda(\theta_2) - \Lambda(\theta_2^-)$, and the complementary slackness condition gives $\tau(\theta_2)(\Lambda(\theta_2) - \Lambda(\theta_2^-)) = 0$: either $\tau(\theta_2) > 0$ and $\Lambda(\theta_2^-) = \Lambda(\theta_2)$, or $\tau(\theta_2) = 0$, in which case $\pi(\theta_2^-) < \pi(\theta_2)$ and δ is continuous at θ_2 (because $\Delta(\theta_2, \pi(\theta_2)) \neq 0$ since $\pi(\theta_2) \neq \pi_f(\theta_2)$), which implies that Λ is continuous at θ_2, so $\Lambda(\theta_2^-) = \Lambda(\theta_2)$. Hence $\underline{\lambda} = \Lambda(\theta_1) = \Lambda(\theta_1^+) = \Lambda(\theta_2^-) = \Lambda(\theta_2) = \bar{\lambda}$.

If $\theta_1 = \underline{\theta}$, then $\underline{\lambda} = \Lambda(\underline{\theta}) = 0$. Similarly, if $\theta_2 = \bar{\theta}$, then $\bar{\lambda} = \Lambda(\bar{\theta}) = 1$.

The last step shows that inequality (12) holds as an equality for $\theta = \theta_1$. Based on (25), it suffices to show that $\delta(\theta_1) = 0$. If $\theta_1 = \underline{\theta}$, the necessary optimality condition (i) in Lemma 2 states that $\delta(\underline{\theta}) = 0$. If $\theta_1 > \underline{\theta}$, condition (ii) in Lemma 2 implies that $\delta(\theta_1) = 0$. To see this, note that $\pi(\theta) < \pi^*$ for $\theta < \theta_1$ because π is non-decreasing and, by assumption $\pi(\theta) \neq \pi^*$ for $\theta < \theta_1$. As a result, condition (ii) in Lemma 2 implies that $\delta(\theta_1) = 0$ if θ_1 is a point of discontinuity of π and $\delta(\theta_1^-) = 0$ if π is strictly increasing over $(\theta_1 - \epsilon, \theta_1]$ for some $\epsilon > 0$. It remains to show that $\delta(\theta_1) = \delta(\theta_1^-)$ if $\pi(\theta_1) = \pi(\theta_1^-)$. If the rule burns money over $(\theta_1 - \epsilon, \theta_1]$, then the complementary slackness condition implies that Λ is constant over $(\theta_1 - \epsilon, \theta_1]$ and $\delta(\theta_1) = \delta(\theta_1^-)$. If the rule does not burn money over $(\theta_1 - \epsilon, \theta_1]$, then because π is strictly increasing, $\Delta(\theta, \pi(\theta)) = 0$ over $(\theta_1 - \epsilon, \theta_1]$ and again, δ is continuous on $(\theta_1 - \epsilon, \theta_1]$. □

A.5 Proof of Proposition 3

Let (π, τ) be an optimal rule. Because π_f is strictly increasing over (θ_1, θ_2), the optimality condition (ii) in Lemma 2 implies that $\delta(\theta) = 0$ for $\theta \in (\theta_1, \theta_2)$ for a non-decreasing Lagrange multiplier function Λ. By linearity of the integral,

$$\int_{(\hat{\theta},\theta_2)} \left[-\nu_\pi(\theta, \pi_f(\theta)) + \rho \frac{\Lambda(\theta) - F(\theta)}{f(\theta)} \right] dF(\theta) = 0$$

for $\hat{\theta} \in (\theta_1, \theta_2)$, where I used that $\Delta(\theta, \pi_f(\theta)) = 0$ for $\theta \in (\theta_1, \theta_2)$. In turn, this implies that

$$\rho\Lambda(\theta) = \rho F(\theta) + \nu_\pi(\theta, \pi_f(\theta)) f(\theta) \tag{26}$$

almost everywhere on (θ_1, θ_2). Because ν_π is continuous, and because Λ is non-decreasing, the condition (26) holds everywhere on (θ_1, θ_2). For $\rho > 0$, the condition that the Lagrange multiplier Λ is non-decreasing implies that $c_d(\theta_1, \theta_2)$ holds.

If $\theta_1 = \underline{\theta}$, because Λ is non-decreasing, $0 = \Lambda(\underline{\theta}) \leq \Lambda(\underline{\theta}^+)$. That is,

$$0 \leq \lim_{\theta \to \theta^+} [\rho F(\theta) + \nu_\pi(\theta, \pi_f(\theta)) f(\theta)] = \nu_\pi(\underline{\theta}, \pi_f(\underline{\theta}^+)) f(\underline{\theta}).$$

Similarly, if $\theta_2 = \bar{\theta}$, because Λ is non-decreasing, $\Lambda(\bar{\theta}^-) \leq \Lambda(\bar{\theta}) = 1$. Hence $\nu_\pi(\bar{\theta}^-, \pi_f(\bar{\theta}^-))f(\bar{\theta}^-) \leq 0$. By continuity and because $f(\bar{\theta}) > 0$, it follows that $\nu_\pi(\bar{\theta}, \pi_f(\bar{\theta})) \leq 0$.

For $\rho = 0$, condition (26) holding everywhere (θ_1, θ_2) implies that $\nu_\pi(\theta, \pi_f(\theta)) = 0$ for $\theta \in (\theta_1, \theta_2)$. □

A.6 Proof of Proposition 4

A.6.1 Necessity

Consider an optimal rule (π, τ) that imposes a graduated schedule of penalties.

An optimal rule is, by definition, admissible and hence incentive compatible. It follows that the money burning schedule necessarily satisfies the envelope condition (4).

Because $\tau(\theta) > 0$ for $\theta > \theta_n$, Proposition 1 implies that $\Delta(\theta, \pi(\theta))$ satisfies condition (10), and condition (11) hold, for $\lambda = 1$ over any sub-interval of $[\theta_n, \bar{\theta}]$ over which π is strictly increasing. Suppose by way of contradiction that $\Delta(\theta, \pi(\theta)) < 0$ for some $\theta \in I$ where $I \subseteq [\theta_n, \bar{\theta}]$ and π is strictly increasing over I. Because π is continuous, there exists $\epsilon > 0$ such that $\Delta(\theta, \pi(\theta)) < 0$ for $\theta \in (\theta - \epsilon, \theta + \epsilon) \cap [\theta_n, \bar{\theta}]$. Incentive compatibility, however, implies that $\tau(\min(\theta + \epsilon, \bar{\theta})) < \tau(\max(\theta - \epsilon, \theta_n))$, which contradicts τ non-decreasing. Hence (c1) holds.

Similarly, because $\tau(\theta) > 0$ for $\theta > \theta_n$, Proposition 2 implies that (c2) holds.

If $\theta_x < \theta_n$, by definition of the rule, Proposition 2 implies that there exists $\underline{\lambda} \leq \bar{\lambda} \in [0, 1]$ such that inequality (12) holds for $\theta \in [\theta_x, \theta_n]$, where $\pi^* = \pi(\theta_x)$ and $\theta_2 = \theta_n$. In addition, the inequality (12) holds as an equality at $\theta = \theta_x$. Because Λ is non-decreasing, $\bar{\lambda} \in [\Lambda(\theta_x), 1]$. The proof of Proposition 2 shows that $\bar{\lambda} = \Lambda(\theta_n)$. In turn, because $\tau(\theta) > 0$ for $\theta > \theta_n$, the complementary slackness condition implies that $\Lambda(\theta_n^+) = 1$. Since Λ is right-continuous on $(\underline{\theta}, \bar{\theta})$, then $\Lambda(\theta_n) = \Lambda(\theta_n^+) = 1$. Hence (c3) holds. For $\underline{\lambda}$, there are two cases. First, if $\theta_x = \underline{\theta}$, then Proposition 2 shows that $\underline{\lambda} = \Lambda(\underline{\theta}) = 0$. If instead, $\underline{\theta} < \theta_x$, then the value of $\underline{\lambda}$ is irrelevant because $\pi(\theta) = \pi(\theta_x)$ by continuity of π. To see this, note that because $\pi(\theta) \leq \pi_f(\theta)$ for $\theta \in [\theta_x, \theta_n]$, by definition of the step function, $\underline{\lambda}$ is irrelevant for inequality (12) for $\theta > \theta_x$, and because $\Delta(\theta_x, \pi(\theta_x)) = 0$, $\underline{\lambda}$ is irrelevant for inequality (12) evaluated at $\theta = \theta_x$.

The proof of Proposition 3 shows that the Lagrange multiplier function satisfies $\Lambda(\theta) = F(\theta) + \frac{1}{\rho}\nu_\pi(\theta, \pi_f(\theta))f(\theta)$, condition $c_d(\underline{\theta}, \theta_x)$ holds and $\nu_\pi(\underline{\theta}, \pi_f(\underline{\theta})) \geq 0$ if $\underline{\theta} < \theta_x$ because $\pi(\theta) = \pi_f(\theta)$ for $\theta \in [\underline{\theta}, \theta_x]$. Hence, (c4) holds.

A.6.2 Sufficiency

Consider a rule (π, τ) that imposes a graduated schedule of penalties parametrized by $\theta_n \in (\underline{\theta}, \bar{\theta})$, $\theta_x \in [\underline{\theta}, \theta_n]$ for which conditions (c1), (c2), (c3), and (c4) hold.

The proof that these conditions are sufficient for the optimality of the rule (π, τ) verifies that the sufficient optimality conditions (i)–(iv) in Lemma 1 hold.

First, let's define the Lagrange multiplier and verify that it is valid. Define the Lagrange multiplier function as follows

$$\Lambda(\theta) = \begin{cases} 1 & \text{for } \theta \in [\theta_x, \bar{\theta}] \setminus \{\underline{\theta}\}, \\ F(\theta) + \frac{1}{\rho}\nu_\pi(\theta, \pi_f(\theta))f(\theta) & \text{for } \theta \in (\underline{\theta}, \theta_x), \\ 0 & \text{for } \theta = \underline{\theta}. \end{cases}$$

The function Λ is non-decreasing over $(\underline{\theta}, \theta_x)$ because, by (c4), condition $c_d(\underline{\theta}, \theta_x)$ holds. If Λ jumps at $\underline{\theta}$, the jump must be positive. If $\underline{\theta} < \theta_x$, then $\nu_\pi(\underline{\theta}, \pi_f(\underline{\theta})) \geq 0$ by (c4), which implies that $\Lambda(\underline{\theta}^+) \geq 0$. If $\underline{\theta} = \theta_x$ instead, then $\Lambda(\underline{\theta}^+) = 1 \geq 0$. If Λ jumps at θ_x, the jump must be positive. If $\theta_x = \theta_n$, then $\pi(\theta_n) = \pi_f(\theta_n)$, which implies that $\Delta(\theta_n, \pi(\theta_n)) = 0$ and, given that the wedge satisfies equality (10) at θ_n for $\lambda = 1$, it follows that Λ does not jump, and hence it is non-decreasing, at θ_n. If $\theta_x \in (\underline{\theta}, \theta_n)$ instead, the condition that inequality (12) holds for $\theta \in [\theta_x, \theta_n]$, where $\pi^* = \pi_f(\theta_x)$, $\underline{\lambda} = 0$, $\bar{\lambda} = 1$, and the inequality (12) holds as an equality at $\theta = \theta_x$ implies that $\Lambda(\theta_x^-) \leq \bar{\lambda}$. To see this, rewrite inequality (12)

$$0 \leq -(1-\rho)\int_\theta^{\theta_n} \Delta(\hat{\theta}, \pi^*) dF(\hat{\theta}) + \int_\theta^{\theta_n} \left[\nu_\pi(\hat{\theta}, \pi^*) - \rho\frac{\bar{\lambda} - F(\hat{\theta})}{f(\hat{\theta})}\right] dF(\hat{\theta})$$

and because inequality (12) holds as an equality for $\theta = \theta_x$, the derivative of the right-hand side must be non-negative at θ_x; that is

$$0 \leq (1-\rho)\Delta(\theta_x, \pi^*)f(\theta_x) - \left[\nu_\pi(\theta_x, \pi^*) - \rho\frac{\bar{\lambda} - F(\theta_x)}{f(\theta_x)}\right] f(\theta_x).$$

Using that $\Delta(\theta_x, \pi^*) = 0$ because $\pi^* = \pi_f(\theta_x)$, the inequality can be rewritten as follows:

$$\Lambda(\theta_x^-) = F(\theta_x) + \frac{1}{\rho}\nu_\pi(\theta_x, \pi_f(\theta_x))f(\theta_x) \leq \bar{\lambda} = 1.$$

Lastly, if $\theta_x = \underline{\theta}$, then the Lagrange multiplier function is non-decreasing because it jumps from 0 to 1 at $\underline{\theta}$ and remains constant at 1 for $\theta \geq \underline{\theta}$.

The function Λ is right-continuous over $(\underline{\theta}, \bar{\theta})$ because ν_π, F, f, and π_f are continuous.

To show that conditions (i) and (ii) in Lemma 1 hold, I show that the equivalent conditions (i) and (ii) in Lemma 2 are implied by conditions (c1), (c2), (c3), (c4), and the envelope condition (4).

Because π is continuous over $(\theta_n, \bar{\theta}]$ by the definition of a graduated schedule of penalties, there exists a countable collection of open intervals (θ_j, θ_{j+1}) indexed by $j \in J$ such that π is constant over (θ_j, θ_{j+1}), and π is strictly increasing over $C \equiv (\theta_n, \bar{\theta}] \setminus \cup_{j \in J}(\theta_j, \theta_{j+1})$. Condition (c1) implies that $\delta(\theta) = 0$ for $\theta \in C$. Condition (c2) implies that at endpoints θ_j of a bunching interval $\delta(\theta_j) = 0$, and $\delta(\theta) \leq \delta(\theta_{j+1}) = 0$ for $\theta \in (\theta_j, \theta_{j+1})$. Also, $\delta(\theta_n) = 0$ because θ_n is either reached from a marginal-deterrence region, so that $\delta(\theta_n) = 0$ by (c1) and the continuity of δ, or the endpoint of a bunching interval, so that $\delta(\theta_n) = 0$ by (c2).

For $\theta \in [\theta_x, \theta_n)$, rewriting δ as in (25) shows that (c3), which assumes that inequality (12) holds for $\theta \in [\theta_x, \theta_n]$, where $\pi^* = \pi(\theta_x)$, $\underline{\lambda} = 0$, $\bar{\lambda} = 1$, and $\theta_2 = \theta_n$, implies that $\delta(\theta) \leq 0$. In addition, because (c3) assumes that inequality (12) holds as an equality at $\theta = \theta_x$, $\delta(\theta_x) = 0$.

For $\theta \in [\underline{\theta}, \theta_x)$, because $\delta(\theta_x) = 0$, $\delta(\theta) = \delta(\theta) - \delta(\theta_x)$. By linearity of the integral, and by definition of Λ over $(\underline{\theta}, \theta_x)$, it follows that $\delta(\theta) = 0$. In turn, because $\pi(\theta) = \pi_f(\theta)$ for $[\underline{\theta}, \theta_x)$, it follows that $\Delta(\theta, \pi(\theta)) = \Delta(\theta, \pi_f(\theta)) = 0$ for $\theta \in [\underline{\theta}, \theta_x)$ and as a result δ is continuous at $\underline{\theta}$ and

$$\delta(\underline{\theta}) = \delta(\underline{\theta}^-) = 0.$$

If $\underline{\theta} = \theta_x$, then $\delta(\underline{\theta}) = 0$ because $\delta(\theta_x) = 0$ as shown above.

The complementary slackness condition (iii) in Lemma 1 holds because $\tau(\theta) = 0$ for $\theta \leq \theta_n$ and $d\Lambda(\theta) = 0$ for $\theta > \theta_x$ and $\theta_x \leq \theta_n$.

It remains to show that the rule is admissible (i.e., condition (iv) in Lemma 1 holds). The money-burning constraint is satisfied because $\tau(\theta) = 0$ for $\theta \in [\underline{\theta}, \theta_n]$ and τ is non-decreasing. To see this, note that the money burning schedule satisfies $\tau(\theta) = 0$ for $\theta \leq \theta_n$ because the policy schedule associated with a graduated schedule of penalties implements the flexible policy for $[\underline{\theta}, \theta_x)$ and there is bunching for $[\theta_x, \theta_n]$. The condition (c1) that marginal deterrence implements a non-negative wedge implies that $\tau(\theta) \geq 0$. The rule is incentive compatible because τ satisfies the envelope condition (4). □

OA1 Upper bound on expected penalties

This section elaborates on the observation that solving for an optimal rule can be instrumental in solving mechanisms with an upper bound on expected money burning.

$$\max_{\pi,\tau} \left\{ \int_\Theta \left[U^P(\theta, \pi(\theta))\right] dF(\theta) \;\middle|\; (2), (3), \text{and} \quad \int_\Theta \tau(\theta) dF(\theta) \leq \bar{\tau} \right\} \tag{27}$$

If expected money burning is non-positive (i.e., $\bar{\tau} = 0$), program (27) corresponds to a delegation problem without the possibility to burn money.

Proposition 5 (Cost of enforcement and expected money burning)**.** *If (π, τ) is an optimal rule then (π, τ) solves (27) for $\bar{\tau} = \int_\Theta \tau(\theta) dF(\theta)$.*

Proof of Proposition 5. Suppose that (π, τ) is an optimal rule for some $\rho \geq 0$ and let $\bar{\tau} = \int_\Theta \tau(\theta) dF(\theta)$. Suppose, by way of contradiction, that (π, τ) is not a solution to (27). That is, there exists $(\hat{\pi}, \hat{\tau})$ that satisfies constraints (2), (3), and $\int_\Theta \hat{\tau}(\theta) dF(\theta) \leq \bar{\tau}$, and

$$\int_\Theta \left[U^P(\theta, \pi(\theta))\right] dF(\theta) < \int_\Theta \left[U^P(\theta, \hat{\pi}(\theta))\right] dF(\theta).$$

This contradicts that (π, τ) is an optimal rule because it satisfies constraints (2) and (3), and

$$\int_\Theta \left[U^P(\theta, \pi(\theta)) - \rho\tau(\theta)\right] dF(\theta) = \int_\Theta \left[U^P(\theta, \pi(\theta))\right] dF(\theta) - \rho\bar{\tau} < \int_\Theta \left[U^P(\theta, \hat{\pi}(\theta)) - \rho\hat{\tau}(\theta)\right] dF(\theta).$$

□

Relatedly, Ambrus and Egorov (2017) study an environment with money burning and an endogenous upper bound on expected money burning. The upper bound is implied by an ex-ante participation constraint and a limit on lump-sum transfers from the principal to the agent. In that environment, the degree of asymmetry in the cost of burning money obtains from the different sensitivity of the payoff of the principal and the agent to the policy. To facilitate the comparison with the analysis in Ambrus and Egorov (2017), note that ρ in this paper corresponds to $1/A$ in Ambrus and Egorov (2017) for quadratic preferences $U^P(\theta, \pi) = -A(\pi - \theta)^2$ and $U^A(\theta, \pi) = -(\pi - \theta - \bar{\nu})^2$.

OA2 Additional proofs

OA2.1 Proof of Lemma (Incentive compatible allocations)

The proof follows the argument in Myerson (1981). Suppose that $\pi(\cdot)$ is incentive compatible given a money-burning schedule $\tau(\cdot)$. Define $V(\theta) = \theta\pi(\theta) + b(\pi(\theta)) - \tau(\theta)$. Consider $\theta > \hat{\theta}$, incentive compatibility implies,

$$V(\theta) - (\theta - \hat{\theta})\pi(\hat{\theta}) \geq V(\hat{\theta}), \quad \text{and} \quad V(\hat{\theta}) - (\hat{\theta} - \theta)\pi(\theta) \geq V(\theta).$$

The inequalities combined imply that $\pi(\cdot)$ is non-decreasing,

$$\pi(\hat{\theta}) \leq \frac{V(\hat{\theta}) - V(\theta)}{\hat{\theta} - \theta} \leq \pi(\theta).$$

Since $\pi(\cdot)$ is non-decreasing, $V(\cdot)$ is continuous and differentiable almost everywhere. Taking the limit, $V'(\theta) = \pi(\theta)$. Integrating from $\underline{\theta}$ to θ gives $V(\theta) = V(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \pi(\theta)d\theta$. Replacing V by its definition gives (4).

Suppose instead that $\pi(\cdot)$ is non-decreasing and, for a given $\tau(\underline{\theta})$, define $\tau(\cdot)$ according to (4). Using the definitions of V, rewrite (4) as follows: $V(\theta) = V(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \pi(\theta)d\theta$. For $\theta \geq \hat{\theta}$,

$$V(\theta) - V(\hat{\theta}) = \int_{\underline{\theta}}^{\theta} \pi(\theta)d\theta - \int_{\underline{\theta}}^{\hat{\theta}} \pi(\theta)d\theta = \int_{\hat{\theta}}^{\theta} \pi(\theta)d\theta \geq \int_{\hat{\theta}}^{\theta} \pi(\hat{\theta})d\theta = (\theta - \hat{\theta})\pi(\hat{\theta}).$$

The inequality holds because $\pi(\cdot)$ is non-decreasing so $\pi(\theta) \geq \pi(\hat{\theta})$. Substituting the definition of V gives the incentive compatibility constraints.

OA2.2 Proof of Corollary 1 and Corollary 2

Proof of Corollary 1. If the fiscal rule leaves some discretion below a cap, i.e., $\theta_p \in (\underline{\theta}, \bar{\theta})$, then Proposition 2 implies that there exists $\underline{\lambda} \in [0, 1]$ such that condition $c_b(\theta_p, \bar{\theta}, \pi(\theta_p), \underline{\lambda}, \bar{\lambda} = 1)$ holds. Because $\pi(\theta_p) \leq \pi_f(\theta_p)$, the wedge $\Delta(\theta_p, \pi(\theta_p)) \geq 0$, and by definition of the step function in condition c_b and inequality (12), $c_b(\theta_p, \bar{\theta}, \pi(\theta_p), \underline{\lambda} = 1, \bar{\lambda} = 1)$ must hold.

If the fiscal rule leaves no discretion, then Proposition 2 implies that condition $c_b(\underline{\theta}, \bar{\theta}, \pi^{ER}, \underline{\lambda} = 0, \bar{\lambda} = 1)$ holds.

Substituting $\nu(\theta, \pi) = (1-\beta)\theta\pi$ in inequality (12) and rearranging terms gives the inequalities in Corollary 1 and concludes the proof. □

Proof of Corollary 2. Proposition 2 implies that condition $c_b(\underline{\theta}, \theta_x, \pi(\theta_x), \underline{\lambda} = 0, \bar{\lambda} = 1)$ holds. Substituting $\nu(\theta, \pi) = (1 - \beta)\theta\pi$ in inequality (12) and rearranging terms gives the inequality in Corollary 2 and concludes the proof. □

OA2.3 Proof of Corollary 3

Because $\pi(\theta) = \pi_f(\theta)$ for $\theta \in (\theta_l, \theta_p)$ and (π, τ) is an optimal rule, Proposition 3 implies that $c_d(\theta_l, \theta_p)$ holds; which implies that for $\theta \in [\theta_l, \theta_p]$

$$\nu_\pi(\theta_l, \pi_f(\theta_l))f(\theta_l) + \rho F(\theta_l) \leq \nu_\pi(\theta, \pi_f(\theta))f(\theta) + \rho F(\theta) \leq \nu_\pi(\theta_p, \pi_f(\theta_p))f(\theta_p) + \rho F(\theta_p).$$

To prove that the right inequality in Corollary 3 holds, first, consider the case $\theta_p = \bar{\theta}$. Then for $\theta \in [\theta_l, \bar{\theta}]$

$$\nu_\pi(\theta, \pi_f(\theta))f(\theta) + \rho F(\theta) \leq \nu_\pi(\bar{\theta}, \pi_f(\bar{\theta}))f(\bar{\theta}) + \rho F(\bar{\theta}) \leq \rho,$$

where the second inequality follows because $\nu_\pi(\bar{\theta}, \pi_f(\bar{\theta})) \leq 0$ as shown in Proposition 3. Rearranging terms gives the right inequality in Corollary 3. Consider the case $\theta_p < \bar{\theta}$ instead. There are two cases to consider because there can either be marginal deterrence or bunching for $\theta > \theta_p$. In either case there exists $\pi^* \geq \pi_f(\theta_p)$ satisfying

$$\nu_\pi(\theta_p, \pi^*) - (1-\rho)\Delta(\theta_p, \pi^*) \leq \rho \frac{\lambda - F(\theta_p)}{f(\theta_p)}. \tag{28}$$

If there is marginal deterrence, then for some $\epsilon > 0$ condition $c_m(\theta_p, \theta_p + \epsilon, \pi, \lambda)$ holds: there exists a constant $\lambda \in [0,1]$ such that $(1-\rho)\Delta(\theta, \pi(\theta)) = \nu_\pi(\theta, \pi(\theta)) - \rho\frac{\lambda - F(\theta)}{f(\theta)}$ for $\theta \in (\theta_p, \theta_p + \epsilon)$. Because π is right-continuous, taking $\theta \downarrow \theta_p$ gives this relation at θ_p, which is (28) with equality and $\pi^* = \pi(\theta_p) \geq \pi_f(\theta_p)$ (an upward jump may occur at θ_p).

If there is bunching, Proposition 2 implies that there exist $\lambda \in [0,1]$, $\pi^* \geq \pi_f(\theta_p)$, and $\hat{\theta} > \theta_p$ such that condition $c_b(\theta_p, \hat{\theta}, \pi^*, \lambda, \lambda)$ holds (with $\underline{\lambda} = \bar{\lambda} = \lambda$ because either $\pi^* = \pi_f(\theta_p)$, in which case the value of $\underline{\lambda}$ does not matter, or $\pi^* > \pi_f(\theta_p)$, in which case bunching is enforced by money burning). Because inequality (12) holds for $\theta \in [\theta_p, \hat{\theta}]$ and as an equality at $\theta = \theta_p$, (28) holds.

The map $\pi \mapsto \rho U^A_\pi(\theta_p, \pi) - U^P_\pi(\theta_p, \pi)$ is non-decreasing (its derivative $\rho U^A_{\pi\pi} - U^P_{\pi\pi} \geq 0$ by $\rho \leq \kappa$) and equals $\nu_\pi(\theta_p, \pi_f(\theta_p))$ at $\pi_f(\theta_p)$ (where $U^A_\pi = 0$) and $\nu_\pi(\theta_p, \pi^*) - (1-\rho)\Delta(\theta_p, \pi^*)$ at π^*. As $\pi^* \geq \pi_f(\theta_p)$, together with (28),

$$\nu_\pi(\theta_p, \pi_f(\theta_p)) \leq \nu_\pi(\theta_p, \pi^*) - (1-\rho)\Delta(\theta_p, \pi^*) \leq \rho \frac{\lambda - F(\theta_p)}{f(\theta_p)}.$$

It follows that, for $\theta \in [\theta_l, \theta_p]$,

$$\begin{aligned}\nu_\pi(\theta, \pi_f(\theta))f(\theta) - \rho(1 - F(\theta)) &\leq \nu_\pi(\theta, \pi_f(\theta))f(\theta) - \rho(\lambda - F(\theta)) \\ &\leq \nu_\pi(\theta_p, \pi_f(\theta_p))f(\theta_p) - \rho(\lambda - F(\theta_p)) \leq 0,\end{aligned}$$

for $\theta \in [\theta_l, \theta_p]$.

A symmetric argument shows that the left inequality in Corollary 3 also holds. □

OA2.4 Proof of Corollary 4

Proof of Corollary 4. The proof uses the necessary and sufficient optimality conditions in Lemma 1 and Lemma 2.

1. *Interval delegation.*

Let $\rho > 0$. To show that (i), (ii), and (iii) in Corollary 4.1 are sufficient optimality conditions, consider the Lagrange multiplier function

$$\Lambda(\theta) = \begin{cases} 1 & \text{for } \theta \in [\theta_p, \bar{\theta}], \\ F(\theta) + \frac{1}{\rho}\nu_\pi(\theta, \pi_f(\theta))f(\theta) & \text{for } \theta \in [\theta_l, \theta_p) \text{ and } \theta \neq \underline{\theta}, \\ 0 & \text{for } \theta \in (\underline{\theta}, \theta_l), \\ 0 & \text{for } \theta = \underline{\theta}. \end{cases}$$

In this proof, I refer to the condition that inequality (12) with $\pi^* = \pi_f(\theta_p)$, $\theta_2 = \bar{\theta}$, and $\underline{\lambda} = \bar{\lambda} = 1$ as inequality (12)$-\theta_p$. Similarly, I refer to the condition that inequality (12) with $\pi^* = \pi_f(\theta_l)$, $\theta_2 = \theta_l$, and $\underline{\lambda} = \bar{\lambda} = 0$ as inequality (12)$-\theta_l$.

The function Λ is non-decreasing over $[\theta_l, \theta_p)$ because condition $c_d(\theta_l, \theta_p)$ holds by condition (ii). It remains to show that the function Λ is non-decreasing at θ_l and θ_p. Note that if $\theta_l > \underline{\theta}$, the jump at θ_l is positive because inequality (12)$-\theta_l$ holds for $\theta \in [\underline{\theta}, \theta_l]$ and as an equality for $\theta = \theta_l$. Hence, the derivative with respect to θ of terms in inequality (12)$-\theta_l$ must satisfy

$$(1-\rho)\Delta(\theta_l, \pi_f(\theta_l))f(\theta_l) \leq \nu_\pi(\theta_l, \pi(\theta_l))f(\theta_l) + \rho F(\theta_l).$$

Because $\Delta(\theta_l, \pi_f(\theta_l)) = 0$, it follows that $\Lambda(\theta_l) \geq 0$. If $\theta_l = \underline{\theta}$, then $\nu_\pi(\underline{\theta}, \pi_f(\underline{\theta})) \geq 0$ and hence $\Lambda(\underline{\theta}^+) \geq 0$. A symmetric argument based on the condition that inequality (12)$-\theta_p$ holds for $\theta \in [\theta_p, \bar{\theta}]$, and as an equality for $\theta = \theta_p$, or $\nu_\pi(\bar{\theta}, \pi_f(\bar{\theta})) \leq 0$ if $\theta_p = \bar{\theta}$ implies that $\Lambda(\theta_p^-) \leq 1$. Also, Λ is right-continuous on $(\underline{\theta}, \bar{\theta})$ because f, ν_π, and π_f are continuous.

Condition (i) in Lemma 2 holds because inequality (12)$-\theta_p$ holds for $\theta \in [\theta_p, \bar{\theta}]$, inequality (12)$-\theta_l$ holds for $\theta \in [\underline{\theta}, \theta_l]$, and the definition of Λ implies that $\delta(\theta) = \delta(\theta_p)$ for $\theta \in (\theta_l, \theta_p)$. To show that $\delta(\underline{\theta}) = 0$ note that (12)$-\theta_l$ holds as an equality for $\theta = \underline{\theta}$ and $\delta(\theta_l) = 0$. If $\theta_l = \underline{\theta}$ instead, then the boundary term in $\delta(\underline{\theta})$ vanishes because $\Delta(\underline{\theta}, \pi(\underline{\theta})) = 0$, so δ is right-continuous at $\underline{\theta}$; since $\delta(\theta) = \delta(\theta_p) = 0$ for $\theta \in (\underline{\theta}, \theta_p)$, it follows that $\delta(\underline{\theta}) = \delta(\underline{\theta}^+) = 0$.

Condition (ii) in Lemma 2 holds because the definition of Λ implies that $\delta(\theta) = \delta(\theta_p)$ for $\theta \in [\theta_l, \theta_p)$ and $\delta(\theta_p) = 0$ (since inequality (12) holds as an equality for $\theta = \theta_p$).

Condition (iii) in Lemma 1 holds because $\tau(\theta) = 0$ for $\theta \in \Theta$.

Condition (iv) in Lemma 1 is satisfied because a rule (π, τ) that delegates an interval is

admissible. The rule (π, τ) is feasible because $\tau(\theta) = 0$ for $\theta \in \Theta$ and it is incentive compatible because the allocation π is non-decreasing and τ satisfies the envelope condition (4).

Conversely, (i) and (iii) in Corollary 4.1 necessarily hold because the rule (π, τ) is optimal and Proposition 2 applies. Condition (ii) in Corollary 4.1 necessarily hold because the rule (π, τ) is optimal and Proposition 3 applies.

For $\rho = 0$, it suffices to show that the flexible allocation and no money burning is optimal if and only if $\nu_\pi(\theta, \pi_f(\theta)) = 0$ for $\theta \in [\underline{\theta}, \bar{\theta}]$. In that case, because $\theta_p = \bar{\theta}$ and $\theta_l = \underline{\theta}$, interval delegation corresponds to $\pi(\theta) = \pi_f(\theta)$ and $\tau(\theta) = 0$ for $\theta \in [\underline{\theta}, \bar{\theta}]$. The necessity part follows from Proposition 3 and the sufficiency part follows from first principles because in the absence of bias the preferred policy of the agent coincides with the policy schedule of the Ramsey rule.

2. *Delegating a narrow mandate.* To show that (i) and (ii) in Corollary 4.2 are sufficient optimality conditions, define Λ so that it coincides with the step function λ in the Proposition 2 with $\theta_1 = \underline{\theta}$, $\theta_2 = \bar{\theta}$, $\underline{\lambda} = 0$ and $\bar{\lambda} = 1$ for $\theta > \underline{\theta}$; that is

$$\Lambda(\theta) = \begin{cases} 1 & \text{if } \theta = \bar{\theta}, \\ 1 & \text{if } \pi^{ER} \leq \pi_f(\theta) \text{ and } \theta \in (\underline{\theta}, \bar{\theta}), \\ 0 & \text{if } \pi^{ER} > \pi_f(\theta) \text{ and } \theta \in (\underline{\theta}, \bar{\theta}), \\ 0 & \text{if } \theta = \underline{\theta}. \end{cases}$$

Since π_f is strictly increasing, Λ is non-decreasing and right-continuous on $(\underline{\theta}, \bar{\theta})$.

In this proof, inequality (12) refers to inequality (12) with $\lambda = \Lambda$ and $\pi^* = \pi^{ER}$, $\theta_2 = \bar{\theta}$, $\theta_1 = \underline{\theta}$.

Condition (i) in Lemma 2 is satisfied because inequality (12) implies that $\delta(\theta) \leq 0$ holds for $\theta \in (\underline{\theta}, \bar{\theta}]$ (see the formula for δ in (25) with $\theta_2 = \bar{\theta}$). Similarly, the condition $\delta(\underline{\theta}) = 0$ holds because inequality (12) holds as an equality for $\theta = \underline{\theta}$.

Condition (ii) in Lemma 2 is trivially satisfied because π is constant.

Condition (iii) in Lemma 1 holds because $\tau(\theta) = 0$ for $\theta \in \Theta$.

Condition (iv) in Lemma 1 holds because the rule (π, τ) is feasible since $\tau(\theta) = 0$ for $\theta \in \Theta$, and it is incentive compatible because the allocation π is non-decreasing and τ satisfies the envelope condition (4).

Conversely, (i) and (ii) in Corollary 4.2 necessarily hold because the rule (π, τ) is optimal and Proposition 2 applies. □

OA2.5 Proof of Lemma 3 on the continuity of rules with asymmetric money burning

Proof. Take π left-continuous, without loss of generality, since the value of π at a jump is welfare-irrelevant; then τ is left-continuous by the envelope condition (4). An optimal rule satisfies the pointwise conditions of Lemma 2 for the associated multiplier Λ. Suppose, by way of contradiction, that π is discontinuous at some interior θ_0 (an endpoint is analogous); since π is non-decreasing, $\pi(\theta_0) = \pi(\theta_0^-) < \pi(\theta_0^+)$, with $\pi(\theta) \leq \pi(\theta_0^-)$ for $\theta < \theta_0$ and $\pi(\theta) \geq \pi(\theta_0^+)$ for $\theta > \theta_0$.

By conditions (i) and (ii), $\delta(\theta_0) = 0$, δ is continuous at θ_0, and $\delta \leq 0$, so θ_0 maximizes δ. Hence, for $\epsilon > 0$,

$$\delta(\theta_0 - \epsilon) - \delta(\theta_0) = \int_{\theta_0-\epsilon}^{\theta_0} A(\theta, \pi(\theta))\, f(\theta)\, d\theta + \rho \int_{(\theta_0-\epsilon,\theta_0]} U_\pi^A(\theta, \pi(\theta))\, d\Lambda \leq 0, \tag{29}$$

$$\delta(\theta_0) - \delta(\theta_0 + \epsilon) = \int_{\theta_0}^{\theta_0+\epsilon} A(\theta, \pi(\theta))\, f(\theta)\, d\theta + \rho \int_{(\theta_0,\theta_0+\epsilon]} U_\pi^A(\theta, \pi(\theta))\, d\Lambda \geq 0, \tag{30}$$

where

$$A(\theta, \pi) := U_\pi^P(\theta, \pi) - \rho U_\pi^A(\theta, \pi) + \rho \tfrac{\Lambda(\theta) - F(\theta)}{f(\theta)}.$$

The proof strategy is to sign the integrals with respect to Λ and then differentiate the remaining, absolutely continuous, integral in ϵ.

The integral with respect to Λ *in* (29) *is non-negative and* $\Lambda(\theta_0^-) = \Lambda(\theta_0)$. If $\pi(\theta_0^-) < \pi_f(\theta_0)$, then for ϵ small $\pi(\theta) \leq \pi(\theta_0^-) < \pi_f(\theta)$ on $(\theta_0 - \epsilon, \theta_0)$, so $U_\pi^A(\theta, \pi(\theta)) > 0$ and the open-interval integral is non-negative; moreover $U_\pi^A(\theta_0, \pi(\theta_0)) = U_\pi^A(\theta_0, \pi(\theta_0^-)) > 0$, so continuity of δ at θ_0 forces $\Lambda(\theta_0) = \Lambda(\theta_0^-)$. If instead $\pi(\theta_0^-) \geq \pi_f(\theta_0)$, then $\pi(\theta_0^+) > \pi(\theta_0^-) \geq \pi_f(\theta_0)$, and strict concavity of $U^A(\theta_0, \cdot)$ gives $U^A(\theta_0, \pi(\theta_0^-)) > U^A(\theta_0, \pi(\theta_0^+))$; by (4), $\tau(\theta_0^-) - \tau(\theta_0^+) > 0$, so $\tau(\theta_0^-) > 0$. By left-continuity $\tau(\theta_0) = \tau(\theta_0^-) > 0$ and $\tau > 0$ on $(\theta_0 - \epsilon, \theta_0)$, so complementary slackness makes Λ constant on $(\theta_0 - \epsilon, \theta_0]$ and the whole integral vanishes.

Thus (29) gives $\int_{\theta_0-\epsilon}^{\theta_0} A(\theta, \pi(\theta)) f(\theta)\, d\theta \leq 0$. Its integrand has a left limit at θ_0, equal to $A(\theta_0, \pi(\theta_0^-)) f(\theta_0)$ since $\Lambda(\theta_0^-) = \Lambda(\theta_0)$, so the right-derivative of the integral at $\epsilon = 0$ is $A(\theta_0, \pi(\theta_0^-)) f(\theta_0)$. As the integral is non-positive and vanishes at $\epsilon = 0$, this derivative is non-positive, so $A(\theta_0, \pi(\theta_0^-)) \leq 0$. Symmetrically, from (30)—with the integral with respect to Λ non-positive and $\Lambda(\theta_0) = \Lambda(\theta_0^+)$ by right-continuity of Λ—we get $A(\theta_0, \pi(\theta_0^+)) \geq 0$.

Finally,

$$\frac{\partial A(\theta_0, \pi)}{\partial \pi} = U_{\pi\pi}^P(\theta_0, \pi) - \rho U_{\pi\pi}^A(\theta_0, \pi) < 0,$$

because $U_{\pi\pi}^P / U_{\pi\pi}^A \geq \kappa > \rho$ and $U_{\pi\pi}^A < 0$. So $A(\theta_0, \cdot)$ is strictly decreasing, and $\pi(\theta_0^-) < \pi(\theta_0^+)$

gives $A(\theta_0, \pi(\theta_0^-)) > A(\theta_0, \pi(\theta_0^+))$, contradicting $A(\theta_0, \pi(\theta_0^-)) \leq 0 \leq A(\theta_0, \pi(\theta_0^+))$. Hence π is continuous. □

OA3 On the enforcement of trade agreements

In this section, I argue that a natural benchmark for the design of a trade agreement enforced by a dynamic use constraint features asymmetric money burning. I show that the cost of enforcement implied by a dynamic use constraint in the repeated game model of trade agreements in Bagwell and Staiger (2005) is asymmetric (i.e., $\rho(\tau) \leq 2/3$ as claimed in footnote 17 in the main text). In the repeated game setting, however, the degree of asymmetry in the cost of money burning is a function of money burning because the benefit to Foreign is a non-linear function of the penalty on Home. In the main text, instead, the degree of asymmetry ρ is assumed to be constant.[28]

The dynamic environment is a repeated static model of trade agreement as in Section 5.4 in Bagwell and Staiger (2005). Consider a two-state dynamic binding system capturing a dynamic use constraint akin to that in the WTO agreement on safeguards. There are two thresholds $\underline{\tau} < \bar{\tau}$. The lower threshold is the best cap that a Ramsey planner would impose on the tariff set by the Home country. Home has the additional flexibility to set a tariff in between $\underline{\tau}$ and $\bar{\tau}$ if it has not done so in the previous period. Intuitively, the cost of exceeding the lower threshold is the forgone flexibility to exceed the threshold again in the next period. The degree of asymmetry in money burning depends on the extent to which Foreign benefits from the limited flexibility for Home to set its tariff.

The objective of this section is to express the benefit to Foreign as a function of the cost to Home of the limited flexibility in its tariff policy. Let π_{cap} denote the profits if the tariffs are set freely below $\underline{\tau}$. Let $\bar{\pi}$ denote the profits associated with the tariff policy $\bar{\tau}$. The welfare of Home if the Home government can set the tariff below $\bar{\tau}$ is $\bar{\nu}$ and

$$(1-\delta)\bar{\nu} = \int_{\underline{\theta}}^{\bar{\theta}} U^A(\theta, \pi_{cap}(\theta))dF(\theta) + \int_{\theta^*}^{\bar{\theta}} [U^A(\theta, \bar{\pi}) - U^A(\theta, \pi_{cap}(\theta)) - \delta(\bar{\nu} - \underline{\nu})]dF(\theta),$$

where θ^* denote the threshold political pressure above which Home activates the escape clause, and $\underline{\nu}$ denotes welfare of Home if the dynamic use constraint binds tariff below $\underline{\tau}$ this period;

[28] The optimal design of a trade agreement in the repeated game setting of Bagwell and Staiger (2005) is left for future research. Athey, Bagwell, and Sanchirico (2001, 2004) show that the solution to the mechanism design problem with money burning is useful to solve for the optimal trade agreement in the repeated game setting. See also Athey, Atkeson, and Kehoe (2005) and Halac and Yared (2025).

that is

$$\underline{\nu} = \int_{\underline{\theta}}^{\bar{\theta}} U^A(\theta, \pi_{cap}(\theta)) dF(\theta) + \delta\bar{\nu}.$$

The cost for Home of exceeding the lower threshold is the forgone flexibility captured by $\bar{\nu} - \underline{\nu}$ in the following period.

Because Home activates the escape clause precisely when it would otherwise be constrained by the cap $\underline{\tau}$, the cap binds for $\theta \geq \theta^*$, so $\pi_{cap}(\theta)$ is constant on the activation region; write $\pi_{cap}(\bar{\theta})$ for this common value. This constancy is what allows $v(\pi_{cap})$ and $U^A(\cdot, \pi_{cap})$ to be pulled out of the tail integrals below.

Similarly, the welfare of Foreign if Home has the flexibility to set tariff below $\bar{\tau}$ is,

$$(1-\delta)\bar{\nu}_F = \int_{\underline{\theta}}^{\bar{\theta}} v(\pi_{cap}(\theta)) dF(\theta) + \int_{\theta^*}^{\bar{\theta}} [v(\bar{\pi}) - v(\pi_{cap}(\theta)) + \delta(\underline{\nu}_F - \bar{\nu}_F)] dF(\theta),$$

where

$$\underline{\nu}_F = \int_{\underline{\theta}}^{\bar{\theta}} v(\pi_{cap}(\theta)) dF(\theta) + \delta\bar{\nu}_F.$$

For the designer of the trade agreement, overall welfare under the two-state dynamic binding system is the sum of Home and Foreign welfare, that is $\bar{w} = \bar{\nu} + \bar{\nu}_F$ and $\underline{w} = \underline{\nu} + \underline{\nu}_F$.

Money burning is asymmetric because Foreign benefits from the forgone flexibility imposed on Home ($\underline{\nu}_F - \bar{\nu}_F > 0$). As a result, the cost of activating the escape clause for Home is higher than it is for the designer of the trade agreement,

$$\begin{aligned}(1-\delta)(\bar{\nu} + \bar{\nu}_F) = &\int_{\underline{\theta}}^{\theta^*} [U^A(\theta, \pi_{cap}(\theta)) + v(\pi_{cap}(\theta))] dF(\theta) \\ &+ \int_{\theta^*}^{\bar{\theta}} [U^A(\theta, \bar{\pi}) + v(\bar{\pi}) - \delta(\bar{\nu} - \underline{\nu}) + \delta(\underline{\nu}_F - \bar{\nu}_F)] dF(\theta).\end{aligned}$$

The degree of asymmetry in the cost of burning money is $1 - \rho(\bar{\tau}) = \frac{\underline{\nu}_F - \bar{\nu}_F}{\bar{\nu} - \underline{\nu}}$. It is a function of $\bar{\tau}$ because $\bar{\nu}_F$ depends on the profits $\bar{\pi}$ associated with $\bar{\tau}$.

To evaluate the degree of asymmetry in the cost of burning money, I express the benefit to Foreign $\underline{\nu}_F - \bar{\nu}_F$ as a function of the cost to Home $\bar{\nu} - \underline{\nu}$. First, I rewrite $\underline{\nu}_F - \bar{\nu}_F$ in terms of profits in the Home country. Substitute

$$\int_{\underline{\theta}}^{\bar{\theta}} v(\pi_{cap}(\theta)) dF(\theta) = \underline{\nu}_F - \delta\bar{\nu}_F$$

into the welfare of Foreign when Home has access to escape clauses, which gives

$$(1-\delta)\bar{\nu}_F = \underline{\nu}_F - \delta\bar{\nu}_F + \int_{\theta^*}^{\bar{\theta}} [v(\bar{\pi}) - v(\pi_{cap}(\theta)) - \delta(\bar{\nu}_F - \underline{\nu}_F)] dF(\theta).$$

After rearranging terms,

$$\underline{\nu}_F - \bar{\nu}_F = [v(\pi_{cap}(\bar{\theta})) - v(\bar{\pi})]\frac{1 - F(\theta^*)}{1 + \delta(1 - F(\theta^*))}$$

Substituting the formula for v and rearranging terms gives

$$\underline{\nu}_F - \bar{\nu}_F = \left[\frac{3}{2}\left(\sqrt{\bar{\pi}} - \sqrt{\pi_{cap}(\bar{\theta})}\right) - \frac{9}{4}(\bar{\pi} - \pi_{cap}(\bar{\theta}))\right]\frac{1 - F(\theta^*)}{1 + \delta(1 - F(\theta^*))}.$$

Similarly, I rewrite $\bar{\nu} - \underline{\nu}$ in terms of profits in the Home country as follows

$$\begin{aligned}\bar{\nu} - \underline{\nu} = 3&\left[\frac{3}{2}\left(\sqrt{\bar{\pi}} - \sqrt{\pi_{cap}(\bar{\theta})}\right) - \frac{9}{4}(\bar{\pi} - \pi_{cap}(\bar{\theta}))\right]\frac{1 - F(\theta^*)}{1 + \delta(1 - F(\theta^*))} \\ &- \left(\frac{7}{4} - E[\theta|\theta \geq \theta^*]\right)(\bar{\pi} - \pi_{cap}(\bar{\theta}))\frac{1 - F(\theta^*)}{1 + \delta(1 - F(\theta^*))}.\end{aligned}$$

It follows that

$$\underline{\nu}_F - \bar{\nu}_F = \frac{1}{3}(\bar{\nu} - \underline{\nu}) + \frac{1}{3}\left(\frac{7}{4} - E[\theta|\theta \geq \theta^*]\right)(\bar{\pi} - \pi_{cap}(\bar{\theta}))\frac{1 - F(\theta^*)}{1 + \delta(1 - F(\theta^*))}.$$

As a result, $\rho(\bar{\tau}) \leq 2/3$ because

$$\rho(\bar{\tau}) = 1 - \frac{\underline{\nu}_F - \bar{\nu}_F}{\bar{\nu} - \underline{\nu}} = \frac{2}{3} - \frac{1}{3}\frac{\left(\frac{7}{4} - E[\theta|\theta \geq \theta^*]\right)(\bar{\pi} - \pi_{cap}(\theta^*))}{\bar{\nu} - \underline{\nu}}\frac{1 - F(\theta^*)}{1 + \delta(1 - F(\theta^*))},$$

and $\bar{\theta} < 7/4$, $\pi_{cap}(\theta^*) < \bar{\pi}$, and $\bar{\nu} - \underline{\nu} > 0$.

OA4 Proofs of the corollaries in the applications

Proof of Corollary 5. Because (π, τ) is optimal and $\pi(\theta) = \pi_f(\theta_1) \equiv \pi^*$ for $\theta \in (\theta_1, \theta_2)$ with $\pi(\theta) \neq \pi^*$ for $\theta \notin [\theta_1, \theta_2]$, Proposition 2 implies that condition $c_b(\theta_1, \theta_2, \pi^*, \underline{\lambda}, \bar{\lambda})$ holds for some $\underline{\lambda} \leq \bar{\lambda}$ in $[0, 1]$. At a tariff binding $\pi^* = \pi_f(\theta_1) \leq \pi_f(\theta)$ for $\theta \geq \theta_1$, so the step function in c_b equals $\bar{\lambda}$ throughout (θ_1, θ_2); write $\lambda \equiv \bar{\lambda}$ (the value $\underline{\lambda}$ multiplies $\Delta(\theta_1, \pi^*) = 0$ and is irrelevant). Rearranging c_b with $U_\pi^P = \Delta - \nu_\pi$ and integrating by parts gives

$$\int_\theta^{\theta_2} U_\pi^P(\hat{\theta}, \pi^*)\, dF(\hat{\theta}) \leq \rho\,\Delta(\theta, \pi^*)(\lambda - F(\theta)) - \rho\,\Delta(\theta_2, \pi^*)(\lambda - F(\theta_2))$$

for $\theta \in [\theta_1, \theta_2]$, with equality at $\theta = \theta_1$. Since $\Delta(\theta_1, \pi^*) = 0$, $\Delta(\theta, \pi^*) = \theta - \theta_1$ on $[\theta_1, \theta_2]$. Using the equality at $\theta = \theta_1$ to substitute $-\rho\,\Delta(\theta_2, \pi^*)(\lambda - F(\theta_2)) = \int_{\theta_1}^{\theta_2} U_\pi^P(\hat{\theta}, \pi^*)\, dF(\hat{\theta})$ gives, for $\theta \in (\theta_1, \theta_2)$,

$$0 \leq \rho(\theta - \theta_1)(\lambda - F(\theta)) + \int_{\theta_1}^{\theta} U_\pi^P(\hat{\theta}, \pi^*)\, dF(\hat{\theta}).$$

For the additive bias of the trade model, $\nu(\theta,\pi) = -v(\pi)$, so $U_\pi^P(\hat\theta,\pi^*) = (\hat\theta-\theta_1)+v'(\pi^*)$. Substituting and integrating,

$$0 \le \rho(\theta-\theta_1)(\lambda-F(\theta)) + v'(\pi^*)\big(F(\theta)-F(\theta_1)\big) + \int_{\theta_1}^{\theta}(\hat\theta-\theta_1)\,dF(\hat\theta).$$

Dividing by $F(\theta)-F(\theta_1)>0$ and rearranging yields

$$\theta_1 - E[\hat\theta \mid \theta_1 \le \hat\theta \le \theta] \le v'(\pi_f(\theta_1)) + \rho(\theta-\theta_1)\frac{\lambda-F(\theta)}{F(\theta)-F(\theta_1)},$$

for $\theta\in(\theta_1,\theta_2)$, with equality at $\theta=\theta_2$. Finally, if $\theta_2=\bar\theta$, then $\bar\lambda=1$ by Proposition 2. If instead $\tau(\theta)>0$ for $\theta>\theta_2$, then the money-burning constraint is slack on $(\theta_2,\bar\theta]$, so Λ is constant there and equal to $\Lambda(\bar\theta)=1$; hence $\lambda=\bar\lambda=1$. □

Proof of Corollary 6. Because the optimal trade agreement grants discretion below $\pi_f(\theta_1)$, i.e. $\pi(\theta)=\pi_f(\theta)$ for $\theta\le\theta_1$, Proposition 3 implies that condition c_d holds: the map $\theta\mapsto\rho F(\theta)+\nu_\pi(\theta,\pi_f(\theta))f(\theta)$ is non-decreasing for $\theta\le\theta_1$. In the trade model $\nu(\theta,\pi)=-v(\pi)$, and evaluating the bias at the flexible policy gives $\nu_\pi(\theta,\pi_f(\theta)) = -v'(\pi_f(\theta)) = \frac{7-4\theta}{12}$. Hence $\rho F(\theta)+\frac{7-4\theta}{12}f(\theta)$ is non-decreasing. Differentiating (for differentiable f) and rearranging,

$$\left(\tfrac{7}{4}-\theta\right)\frac{f'(\theta)}{f(\theta)} \ge 1-3\rho, \quad \theta\le\theta_1.$$ □

Proof of Corollary 7. Because the trade agreement burns money on an interval over which the policy is strictly increasing, Proposition 1 implies that, over $[\theta_1,\theta_2]$, the marginal penalty induces the wedge (10) and the Lagrange multiplier on the money-burning constraint satisfies the monotonicity condition

$$\rho\frac{d}{d\theta}\frac{\lambda-F(\theta)}{f(\theta)} \ge \nu_{\pi\theta} - (1-\rho).$$

The trade-model bias $\nu(\theta,\pi)=-v(\pi)$ is independent of θ, so $\nu_{\pi\theta}=0$ and the monotonicity condition reduces to $\rho\frac{d}{d\theta}\frac{\lambda-F(\theta)}{f(\theta)} \ge \rho-1$, with $\nu_\pi(\theta,\pi(\theta)) = -v'(\pi(\theta))$ in (10). Finally, if $\tau(\theta)>0$ for $\theta>\theta_1$ then the money-burning constraint is slack on a right-neighborhood, so the multiplier is at its upper value and $\lambda=1$. □

Proof of Corollary 8. Throughout, F is uniform on $[\underline\theta,\bar\theta]$, so $f=1/(\bar\theta-\underline\theta)$ and the inverse hazard rate is $\frac{1-F(\theta)}{f(\theta)} = \bar\theta-\theta$. Along the flexible policy the bias is $\nu_\pi(\theta,\pi_f(\theta)) = -v'(\pi_f(\theta)) = \frac{7-4\theta}{12}$, and the wedge $\Delta(\theta,\pi_f(\theta))=0$. The maintained assumptions are $\bar\theta<7/4$, $\bar\theta-\underline\theta>2(\frac{7}{4}-\bar\theta)$, and $\rho<\kappa=2/3$. Define

$$\rho_1=\tfrac{1}{2}, \qquad \rho_2 = \frac{7-4\underline\theta}{12(\bar\theta-\underline\theta)} = \tfrac{1}{2} - \frac{3\bar\theta-\frac{7}{2}-\underline\theta}{6(\bar\theta-\underline\theta)}.$$

Because $\bar\theta<7/4$ and $\bar\theta-\underline\theta>2(\frac{7}{4}-\bar\theta)$, one has $1/3<\rho_2<\rho_1$.

Step 1: the necessary conditions split the line at $\rho_1 = \frac{1}{2}$. For uniform F, the inverse hazard rate has slope -1. Corollary 7 then requires, for marginal deterrence over an interval, $\rho \frac{d}{d\theta}\frac{1-F}{f} = -\rho \geq \rho - 1$, i.e. $\rho \leq \frac{1}{2}$. Conversely, Corollary 5 specialized to uniform F (with $E[\hat{\theta} \mid \theta_1 \leq \hat{\theta} \leq \theta] = \frac{\theta_1+\theta}{2}$ and $\frac{1-F(\theta)}{F(\theta)-F(\theta_1)} = \frac{\bar{\theta}-\theta}{\theta-\theta_1}$) reduces to $(\rho - \frac{1}{2})\theta \leq v'(\pi_f(\theta_1)) + \rho\bar{\theta} - \frac{\theta_1}{2}$ for all θ up to the binding, which can hold up to $\bar{\theta}$ only if $\rho \geq \frac{1}{2}$. Hence for $\rho > \frac{1}{2}$ an optimal agreement cannot feature marginal deterrence and disciplines only by limiting the choice set, whereas for $\rho < \frac{1}{2}$ it cannot feature bunching at a binding and must discipline through marginal deterrence.

Step 2: part a) ($\rho \geq \rho_1$). By Step 1 there is no marginal deterrence, so the optimal agreement is an interval-delegation rule: discretion below a cap at $\pi_f(\theta_p)$ and bunching above. The cap binds at the flexible policy, so $\Delta(\theta_p, \pi_f(\theta_p)) = 0$ and the equality at $\theta = \theta_p$ in condition $c_b(\theta_p, \bar{\theta}, \pi_f(\theta_p), 1, 1)$ of Proposition 2 reduces to the cap equality,

$$-v'(\pi_f(\theta_p)) = E[\theta \mid \theta \geq \theta_p] - \theta_p$$

which is equivalent to $\theta_p = 3\bar{\theta} - \frac{7}{2}$. The range assumption gives $\theta_p > \underline{\theta}$, so the cap leaves discretion below the binding (tariff overhang). The conditions (i)–(iii) of Corollary 4 hold for uniform F and $\rho \geq \frac{1}{2}$ (the cap inequality is the $\theta_2 = \bar{\theta}$, $\lambda = 1$ specialization of Step 1, satisfied for $\rho \geq \frac{1}{2}$; condition (iii) holds because the mean residual life of the uniform distribution is decreasing), so the cap is optimal.

Step 3: parts b) and c) ($\rho < \rho_1$). By Step 1 the optimal agreement disciplines through marginal deterrence wherever it is active (with $\lambda = 1$, since money burning is positive on a right-neighborhood of any point at which it occurs). Substituting $\Delta(\theta, \pi) = \theta + b'(\pi)$ and $\nu_\pi(\theta, \pi) = -v'(\pi)$ into the wedge equation $(1-\rho)\Delta(\theta, \pi_n(\theta)) = \nu_\pi(\theta, \pi_n(\theta)) - \rho(\bar{\theta} - \theta)$ and solving for $\pi_n(\theta)$ shows that $\Delta_n(\theta) \equiv \Delta(\theta, \pi_n(\theta))$ is affine in θ, with slope $\frac{3\rho-1}{2-3\rho}$; because $\rho < \kappa = 2/3$, the denominator is strictly positive, so the sign of the slope matches the sign of $3\rho - 1$.

The threshold θ_n at which money burning begins with a continuous (differentiable) penalty schedule is the root of Δ_n, obtained from $\pi_n(\theta_n) = \pi_f(\theta_n)$ that holds at any such root:

$$\tfrac{7-4\theta_n}{12} = \rho(\bar{\theta} - \theta_n),$$

which is equivalent to $\theta_n(\rho) = \frac{7-12\rho\bar{\theta}}{4(1-3\rho)}$. On each of $(0, \frac{1}{3})$ and $(\frac{1}{3}, \frac{1}{2})$ separately, $\theta_n(\rho)$ is strictly increasing in ρ (since $d\theta_n/d\rho = 12(7 - 4\bar{\theta})/(4 - 12\rho)^2 > 0$ using $\bar{\theta} < 7/4$), with $\theta_n(0) = 7/4$, $\theta_n(\rho) \to +\infty$ as $\rho \to \frac{1}{3}^-$, $\theta_n(\rho) \to -\infty$ as $\rho \to \frac{1}{3}^+$, and $\theta_n(\frac{1}{2}^-) = 3\bar{\theta} - \frac{7}{2} = \theta_p$. Because Δ_n is affine with root $\theta_n(\rho)$ and slope of sign $\text{sign}(3\rho - 1)$, $\Delta_n(\theta) \geq 0$ iff $\theta \leq \theta_n(\rho)$ when $\rho < \frac{1}{3}$, and iff $\theta \geq \theta_n(\rho)$ when $\rho > \frac{1}{3}$.

For $\rho \in (0, \frac{1}{3})$: since $\theta_n(0) = 7/4 > \bar{\theta}$ and θ_n is increasing, $\theta_n(\rho) > \bar{\theta}$ throughout, so $\Delta_n(\theta) \geq 0$ for every $\theta \in [\underline{\theta}, \bar{\theta}]$. For $\rho \in (\frac{1}{3}, \rho_2)$: since $\theta_n(\rho_2) = \underline{\theta}$ (the defining property of ρ_2, restated below)

and θ_n is increasing on $(\frac{1}{3}, \frac{1}{2})$, $\theta_n(\rho) < \underline{\theta}$, so again $\Delta_n(\theta) \geq 0$ throughout $[\underline{\theta}, \bar{\theta}]$. Hence $\Delta_n(\theta) \geq 0$ on the whole support for every $\rho \in (0, \rho_2)$: $\lambda = 1$ is validated, money burning is non-decreasing, and there is no cap over this range.

For $\rho \in [\rho_2, \rho_1)$, by contrast, $\theta_n(\rho) \in [\underline{\theta}, \theta_p)$ and Δ_n is increasing, so $\Delta_n(\theta) < 0$ for $\theta < \theta_n(\rho)$ and $\Delta_n(\theta) \geq 0$ for $\theta \geq \theta_n(\rho)$: the wedge changes sign within the support, ruling out marginal deterrence below $\theta_n(\rho)$ and calling for discretion there instead.

Recall $\rho_2 = \dfrac{7 - 4\underline{\theta}}{12(\bar{\theta} - \underline{\theta})}$, defined so that $\theta_n(\rho_2) = \underline{\theta}$; because $\bar{\theta} < 7/4$ and $\bar{\theta} - \underline{\theta} > 2(\frac{7}{4} - \bar{\theta})$, $\rho_2 \in (\frac{1}{3}, \frac{1}{2})$.

Part b) ($\rho_2 \leq \rho < \rho_1$). Here $\theta_n(\rho) \in [\underline{\theta}, \theta_p)$, so the candidate is the graduated-penalties allocation with empty bunching region ($\theta_x = \theta_n$): discretion $\pi = \pi_f$ on $[\underline{\theta}, \theta_n)$ and marginal deterrence on $(\theta_n, \bar{\theta}]$, joined with a continuous marginal penalty (no kink, no bunching at the threshold). Condition $c_d(\underline{\theta}, \theta_n)$ holds because $\rho \geq \rho_2 > \frac{1}{3}$ (so the uniform overhang condition $0 \geq 1 - 3\rho$ of Corollary 6 holds), conditions (c1)–(c2) hold by the monotonicity and non-negativity just established, and (c3) is vacuous. Proposition 4 implies the agreement is optimal. There is tariff overhang on $[\underline{\theta}, \theta_n)$ and safeguards above.

Part c) ($0 < \rho < \rho_2$). Here $\Delta_n(\theta) \geq 0$ throughout $[\underline{\theta}, \bar{\theta}]$ (established above, for both $\rho < \frac{1}{3}$ and $\rho \in (\frac{1}{3}, \rho_2)$); in particular the wedge is strictly positive at $\underline{\theta}$. Money burning therefore cannot occur on a right-neighborhood of $\underline{\theta}$: the jump at $\underline{\theta}$ contributes $\rho\Delta(\underline{\theta}, \pi(\underline{\theta}))$ to $\delta(\underline{\theta})$, so a strictly positive wedge at $\underline{\theta}$, with money burning for $\theta > \underline{\theta}$, violates $\delta(\underline{\theta}) = 0$. The smooth marginal-deterrence schedule thus cannot be continued to the bottom, and the candidate is the graduated-penalties allocation with an exemption from the bottom ($\theta_x = \underline{\theta}$): bunching at $\pi(\theta_n)$ on $[\underline{\theta}, \theta_n]$ enforced by a kink in the penalty schedule, and marginal deterrence (differentiable) on $(\theta_n, \bar{\theta}]$, where θ_n solves the exemption equality of condition (c3), $c_b(\underline{\theta}, \theta_n, \pi(\theta_n), 0, 1)$. Conditions (c1)–(c3) hold as in part b), so Proposition 4 implies the agreement is optimal. There is bunching at the threshold and no tariff overhang. □

Proof of Corollary 9. Because (π, τ) is optimal with a price cap at $p(\pi(\theta_p))$, i.e. $\pi(\theta) = \pi(\theta_p)$ for $\theta \geq \theta_p$, Proposition 2 gives condition $c_b(\theta_p, \bar{\theta}, \pi(\theta_p), 1, 1)$. The regulatory bias is $\nu(\theta, \pi) = -\frac{1}{\alpha} v(\pi)$, with $\nu_\pi = -\frac{1}{\alpha} v'$ and $\rho = 1 - \frac{1}{\alpha}$. Because the optimal policy is decreasing in the cost θ, the wedge is $\Delta(\theta, \pi(\theta_p)) = -\theta + \theta_p + \Delta(\theta_p, \pi(\theta_p))$ and the bunching inequality (12) reverses. Using the equality at $\theta = \theta_p$ to eliminate $(1 - \rho)\Delta(\theta_p, \pi(\theta_p))$, and cancelling the constants θ_p inside the integrals exactly as in the proof of Corollary 1, gives

$$\alpha\big(E[\hat{\theta} \mid \hat{\theta} \geq \theta] - \theta\big) + \theta \leq \alpha\big(E[\hat{\theta} \mid \hat{\theta} \geq \theta_p] - \theta_p\big) + \theta_p, \quad \theta \in [\theta_p, \bar{\theta}].$$

The equality at $\theta = \theta_p$ determines the threshold: using $v'(\pi) = p'(\pi)\pi$ and $b'(\pi) = p'(\pi)\pi + p(\pi)$,

$$p(\pi(\theta_p)) = \alpha\big(E[\theta \mid \theta \geq \theta_p] - \theta_p\big) + \theta_p. \qquad \square$$

Proof of Corollary 10. Because the regulation levies taxes (burns money) on an interval $[\theta_1, \theta_2]$ over which the policy is monotone, Proposition 1 implies that the optimal marginal penalty induces the wedge (10) on $[\theta_1, \theta_2]$. Specialize to the regulatory bias $\nu(\theta, \pi) = -\frac{1}{\alpha}v(\pi)$, with $\rho = 1 - \frac{1}{\alpha}$, so that the scaling factor $\frac{1}{1-\rho} = \alpha$. Because output is decreasing in the cost θ, condition (10) applies in the reversed type order. Substituting $\nu_\pi = -\frac{1}{\alpha}v'$ and using $b'(\pi) + v'(\pi) = p(\pi)$ gives the regulated price

$$p(\theta) = \theta + (\alpha - 1)\frac{\lambda - F(\theta)}{f(\theta)}, \qquad \theta \in [\theta_1, \theta_2],$$

while the monotonicity condition (11), with $\nu_{\pi\theta} = 0$, becomes

$$(\alpha - 1)\frac{d}{d\theta}\,\frac{\lambda - F(\theta)}{f(\theta)} \geq -1.$$

If $\tau(\theta) > 0$ for $\theta \geq \theta_1$, the money-burning constraint is slack and the multiplier is at its upper value, so $\lambda = 1$. $\square$

Proof of Corollary 11. Because the regulation grants discretion to implement profits in $[\pi_f(\theta_1), \pi_f(\theta_2)]$, i.e. $\pi(\theta) = \pi_f(\theta)$ on (θ_1, θ_2), Proposition 3 implies that condition $c_d(\theta_1, \theta_2)$ holds: the map $\theta \mapsto \rho F(\theta) + \nu_\pi(\theta, \pi_f(\theta))f(\theta)$ is non-decreasing on (θ_1, θ_2). Specialize to the constant-elasticity regulation, with $\rho = 1 - \frac{1}{\alpha}$ and $\nu_\pi(\theta, \pi_f(\theta)) = \frac{1}{\alpha}\frac{1}{\epsilon}\pi_f(\theta)^{-1/\epsilon}$. Using the flexible price $p(\pi_f(\theta)) = \frac{\epsilon}{\epsilon - 1}\theta$ and the slope $\pi_f'(\theta) = -\epsilon\,\pi_f(\theta)/\theta$, differentiating the non-decreasing map and rearranging yields

$$\theta\frac{f'(\theta)}{f(\theta)} \geq -(\alpha - 1)(\epsilon - 1) - 1, \qquad \theta \in [\theta_1, \theta_2]. \qquad \square$$